VOLUME X OCTOBER, 1931 NUMBER 4

THE BELL SYSTEM TECHNICAL JOURNAL

DEVOTED TO THE SCIENTIFIC AND ENGINEERING ASPECTS OF ELECTRICAL COMMUNICATION

AMERICAN TELEPHONE AND TELEGRAPH COMPANY
NEW YORK

50c per Copy *$1.50 per Year*

THE BELL SYSTEM TECHNICAL JOURNAL

Published quarterly by the
American Telephone and Telegraph Company
195 Broadway, New York, N. Y.

SUBSCRIPTIONS

Subscriptions are accepted at $1.50 per year. Single copies are fifty cents each.
The foreign postage is 35 cents per year or 9 cents per copy.

PRINTED IN U. S. A.

The Bell System Technical Journal

October, 1931

The Interconnection of Telephone Systems—Graded Multiples

By R. I. WILKINSON

The general problem of subscriber interconnection is stated here, while some of the economic and service factors in the selection of trunking systems are briefly considered. The characteristic manner in which telephone calls fall upon ordinary straight trunk groups is presented from both common sense and theoretical standpoints.

One of the widely used trunk rearrangements by which an improved capacity may be achieved under certain conditions is known as graded multiple. A theoretical analysis of this scheme is given, from which are constructed curves for common probabilities of loss. Illustrative examples are included to make clear their use.

A detailed comparison between theory and observation is made with considerable attention paid to critically examining the validity of the assumptions underlying the theory. It is concluded that the present graded multiple engineering tables are based upon a proper modification of the theoretical formula.

Introduction

WITH the completion of the third commercial telephone instrument some fifty odd years ago was born the problem of interconnection. And as the system has grown so has the demand for a universal service. The present complexity of our communication

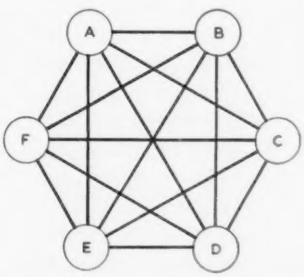


Fig. 1—Direct interconnection of subscribers.

network makes it difficult to appreciate that in those early days it was a comparatively simple thing to provide a direct line from each subscriber to every other subscriber, as shown in the schematic interconnection of the six telephone stations in Fig. 1. In such schemes there

is no possibility of the line being busy; the only faults a subscriber may properly find with the service (over and above transmission troubles) are that "the called party is busy" on another line, and "he does not answer," conditions beyond the control of the interconnecting system.

For a very small number of subscribers all in close proximity such a scheme would and does serve admirably. As soon as the number of subscribers, "n," is increased, however, the number of lines, which equals $\frac{n(n-1)}{2}$, goes up at an enormous rate, almost as the square of the number of subscribers. Since a major element of the cost is in direct variation with the number of lines this plan, even on a modest scale, is quickly prohibited.

Nevertheless, if a truly universal service is to be furnished it must not only be *possible* for any subscriber to communicate directly with any other, but it must be *easily* possible. This problem was solved by the development of the central office plan of interconnection.

Elementary Studies

We may represent a simplified central office exchange system by the line diagram of Fig. 2. Two sets of 100 subscribers, A and B, may make calls to the opposite set via the 10 "trunks" C.[1] A line which connects an individual subscriber to his central office exchange is unique in that it will never be used except when *he* is talking. The lines C, however, which are provided for establishing connections between the telephones in one office and those in the other may well carry calls originated by a large number of subscribers. Thereupon it is readily seen that one interoffice line (or trunk, as we shall hereafter call it) may easily attain a very much higher efficiency, as measured by the per cent time it is in use, than an individual subscriber's line. When one subscriber is not using a particular trunk it is available for use by another; thus we make one trunk do the work for which two or more lines were required in the original arrangement of Fig. 1.

To obtain this increased call carrying capacity per trunk and the consequent savings due to reduction in the total number of trunks, it is necessary to forego one particular advantage: we cannot be absolutely sure that there will be an idle trunk available when each subscriber desires to place his various calls. For it is possible, although very improbable, that all of the subscribers might want to call one another simultaneously, and having far fewer trunks than subscribers in either office many would fail to get immediate service. The

[1] For our purpose it is unnecessary to consider how two subscribers within the set A, or the set B, may be interconnected.

analyses necessary to determine the theoretical probability (or the proportion of times in the long run) that a particular subscriber will find all of a group of trunks busy are well known; and tables and curves are available showing for wide ranges of loads and numbers of trunks the probability of a particular subscriber finding them all in use.[2]

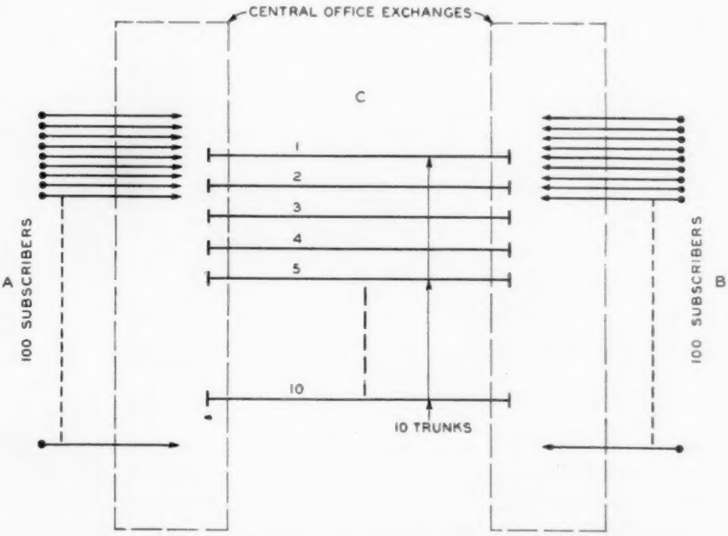


Fig. 2—A simplified central office interconnection system.

In Fig. 3 is shown the load, a, in average simultaneous calls which theoretically may be submitted to c trunks so that, on the average, one one-hundredth ($P = .01$), or one one-thousandth ($P = .001$), of all the calls submitted will find no idle trunk available. By replotting Fig. 3 to show as in Fig. 4 the average load carried per trunk (efficiency) we see that a large group of trunks is relatively much more efficient than a smaller one.

For example, to carry a load of $a = 41$ at $P = .01$ we should provide a single group of 57 trunks, while if we are required to carry the same total load over groups of $c = 16$ trunks, we shall need five such groups or a total of 80 trunks. That this should be so may become clearer by considering, say 20 trunks, first as a complete group and then as two split groups or subgroups of 10 trunks, each carrying one-half of

[2] "The Theory of Probabilities Applied to Telephone Trunking Problems," by E. C. Molina, *Bell System Technical Journal*, November, 1922.

the total load as pictured in Fig. 5. A trunk is represented by each horizontal line, and the load submitted by the vertical arrow, as though it were starting at the bottom of the group to hunt for the first idle trunk. We first observe that the proportion of calls lost on the

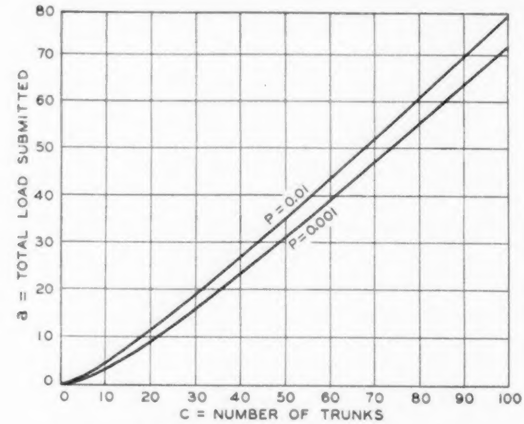


Fig. 3—Load carried by a trunk group at a constant loss.

average in the split groups cannot be less than in the complete group since when all the trunks are full in either case the calls coming in will be delayed or lost altogether, and only when *all 20 trunks* are occupied in the case of the complete group will calls be lost. On the other hand,

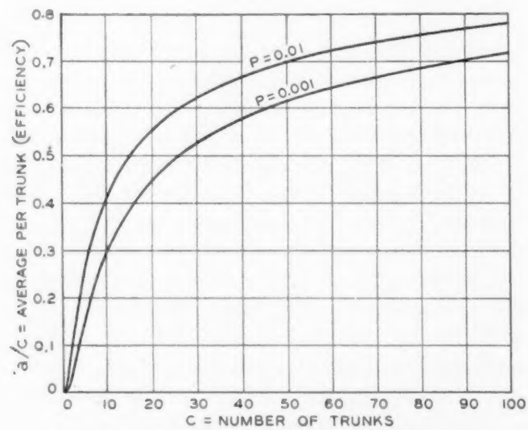


Fig. 4—Load carried per trunk at a constant loss.

in the split group calls may be lost when as few as 10 trunks are busy provided they are all in the group to which the calls at the moment are being originated. Thus the splitting of the group and the consequent reduction of the access may prevent a call in one subgroup, upon finding its 10 trunks busy, from continuing on over the remaining

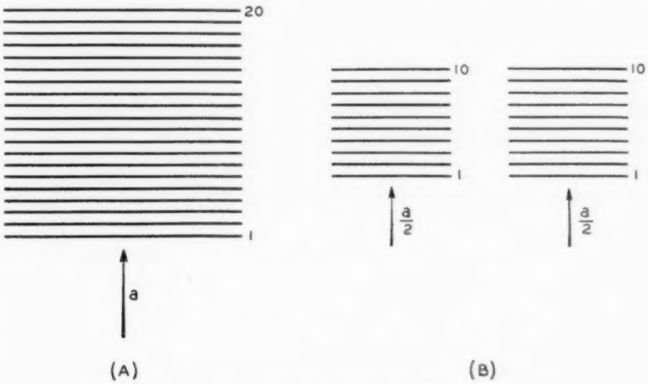


Fig. 5—Comparison of arrangements of twenty trunks.

trunks, one of which might have been idle. We conclude then that, other things being equal, a given load may be more economically carried over one large group than over two or more smaller groups.

Switching Limitations

Unfortunately, "other things" are decidedly not equal. Three major considerations may be pointed out which quite definitely tend to limit the number of trunks to which a particular source may be given access:

1. The high cost of switches having a large number of contacts.
2. The undesirable long hunts which occur in trunk groups of large size.
3. The double connections which increase directly with the load carried.

Of these the first two are usually the more instrumental in regulating the practicable upper limits to the access or hunt. When considered with the efficiency of the trunk groups in a system they comprise, in general, the fundamental data for determining the appropriate arrangement which may be economically employed to handle any given amount of traffic.

We are then faced with the problem of obtaining the maximum

efficiency (maximum average load per trunk) with a given hunt or access. We have seen that the ultimate efficiency is obtained when the total trunks are arranged in a complete or straight grouping. We have also noted how greatly the efficiency is reduced by splitting the total trunks into distinct individual subgroups. It may well be that by certain rearrangements we shall be able to increase this low efficiency without increasing the access.

Graded Multiple Theory

The "graded multiple" means for improving the efficiency of trunking multiples requiring more channels than a single switch can profitably hunt over was proposed in 1905 by E. A. Gray.[3] We may gain an insight into the manner of working of graded multiple by considering a very simple example.

Suppose we have two 10-trunk subgroups as in Fig. 6(A), each carrying an average load corresponding to some predetermined probability of loss, say $P = .001$. Then the approximate average load carried by each individual trunk, *provided all calls hunt over the trunks in the same order*, is shown in Table I.[4] The important point to observe here is that the first trunk is busy a goodly proportion of the time [$a(1) = .748$], the second trunk a somewhat shorter time, and so on down to the last trunks which are comparatively lightly loaded, the tenth trunk being busy only about one-half of one per cent of the time.

The same distribution is approximately maintained in both subgroups of 10 trunks, each of which on the average presents all of its 10 trunks as busy to one out of each thousand calls submitted to it; but it is quite unlikely that this busy condition would occur simultaneously on the two groups. Hence, if on those occasions when a call is being lost in one group it could be allowed to hunt over, say, the last half of the other group, in many cases it would find an idle trunk

[3] E. A. Gray, assignor to the American Telephone and Telegraph Company, filed application July 30, 1907. The patent, No. 1002388, was granted September 5, 1911, for "A Method of and Means for Connecting Telephone Apparatus."

[4] By Erlang's statistical equilibrium method, upon the assumption that "lost" calls are immediately cleared and do not reenter the system, we find the average carried on the rth trunk is

$$a(r) = a[B(r-1, a) - B(r, a)],$$

where $B(x, a)$ is the proportion of traffic passing beyond the xth trunk and may be expressed,

$$B(x, a) = \frac{\dfrac{a^x}{\underline{|x}}}{1 + a + \dfrac{a^2}{\underline{|2}} + \dfrac{a^3}{\underline{|3}} + \cdots + \dfrac{a^x}{\underline{|x}}}.$$

In all cases "a" refers to the average number of simultaneous calls being submitted to an individual set of trunks.

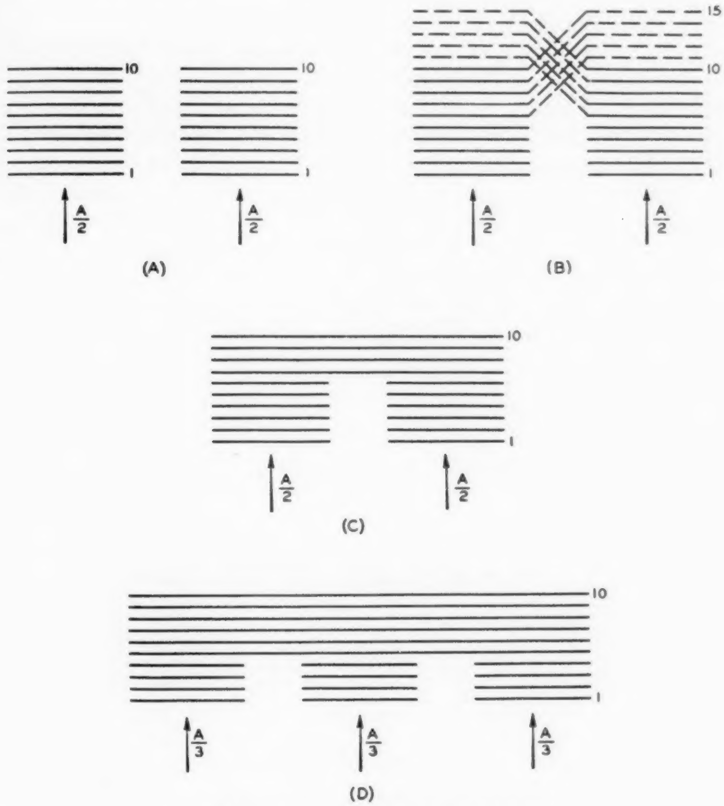


Fig. 6—Genesis of a graded multiple.

available. That is to say, we should be able to capitalize the possibilities of teamwork between the more lightly loaded parts of the groups.

TABLE I

Load Carried by Each Trunk of a Straight Multiple

Average Submitted = a = 2.96

	Number of Trunk in Order Hunted Over									
	1	2	3	4	5	6	7	8	9	10
Load carried on each trunk..	.748	.658	.544	.413	.279	.170	.087	.039	.015	.005

This may be done by arranging the two groups as in Fig. 6(B) so that the cross-connection of the last choice trunks provides an oppor-

tunity for their mutual use by the calls from either subgroup. In this particular case, however, the hunt or access has been increased from 10 to 15 terminals. If we wish to retain the operation of the system on 10-point switching equipment we must compress the trunks into some such form as indicated in Fig. 6(*C*). In so doing we have very likely increased the split group efficiency of the trunks but, at the same time, on account of their fewer total number the load originally submitted may not be adequately served. Hence, a remedy such as shown in Fig. 6(*D*) may perhaps be devised: that is, the addition of more subgroups of the restricted-availability trunks. The study of the actual carrying capacities of these various arrangements is reserved for a later point in this paper.

In general, then, we may represent any such plan of trunking by the schematics of Figs. 7(*A*) and 7(*B*). These are called "simple

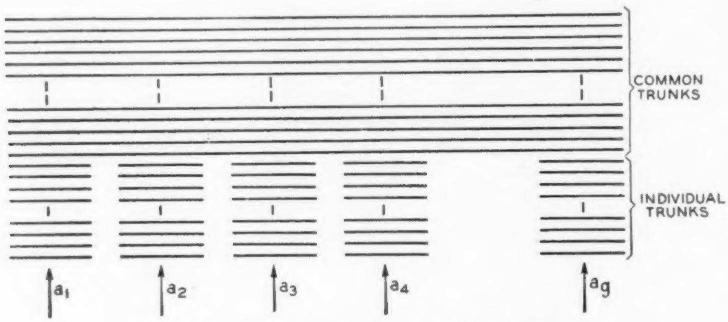


Fig. 7(*A*)—Simple graded multiple with *g* subgroups.

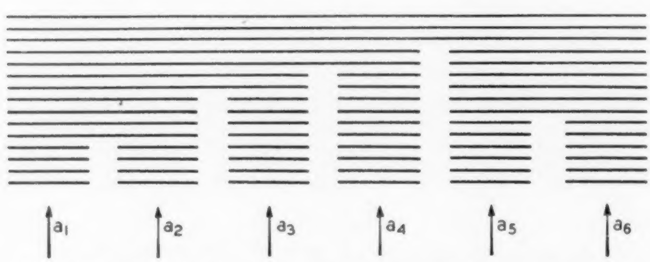


Fig. 7(*B*)—Example of a non-symmetrical progressive graded multiple.

graded multiples" and "progressive graded multiples," respectively, and by varying the number and placement of the trunks composing them a very large variety of arrangements indeed may be obtained. Here all calls hunt from the "bottom" of the group over a particular

one of several sets of "individual" trunks called a subgroup, and if none are idle they continue on up to the "common" or "partially common" trunks which are made accessible to all of the switches that appear before the contributing subgroups. The essential characteristics of the graded multiple are, then: the trunks hunted over first serve a minimum number of switches while those trunks hunted over last are multipled before more or all of the switches. Provided our preliminary analysis is correct, the teamwork obtained between the latter portions of all of the subgroups should result in an increased average per trunk carrying capacity throughout the whole multiple.

Efficiency of Graded Multiples

Several theories for calculating the grade of service of graded multiples with any given submitted load per subgroup have been advanced. All of these involve certain assumptions or empirical approximations.

Dr. Fritz Lubberger [5] of the Siemens and Halske Manufacturing Company, Berlin, in collaboration with Dr. G. Rückle,[6] has presented a universal scheme, partly theoretical, partly empirical, for estimating the load carried by any trunk in a split or graded, or a combination of the two, multiples. That remaining load which is not carried then constitutes the overflow ("Verkehrsreste") from which the proportion "lost" may immediately be determined. The nub of Lubberger's method consists in the application of so-called "Zuschlagsfaktors" to correct the submitted load for the loss when splitting, and the gain when grading. This modified load, on the assumption that it has been reduced to the appropriate equivalent load, is then used to enter a chart constructed for a straight multiple. Combined with this procedure is the assumption that the busy hour loads in each subgroup may not occur simultaneously, thus giving a still freer opportunity for a cooperative usage of the common trunks.

A second plan for estimating the probability of loss of a graded multiple was proposed by the late Dr. M. Merker of Antwerp.[7] He developed a very complicated formula which involves a consideration of the various ways in which a graded multiple might accommodate a given number of calls.

The British Post Office has made some interesting and valuable

[5] F. Lubberger: "Die Wirtschaftlichkeit der Fernsprechanlagen für Ortsverkehr." R. Oldenbourg, Munchen und Berlin, 1927.

[6] G. Rückle und F. Lubberger: "Der Fernsprechverkehr als Massenerscheinung mit starken Schwankungen." Julius Springer, Berlin, 1924.

[7] M. Merker: "Some Notes on the Use of the Probability Theory to Determine the Number of Switches in an Automatic Telephone Exchange." *The Post Office Electrical Engineers' Journal*, vol. 17, Part I, April, 1924.

empirical investigations of the graded multiple problem.[8] These have included successively or progressively-graded multiples as well as those involving only a single set of subgroups feeding into a simple group of common trunks. They conclude that for more than two subgroups the successive grades are somewhat advantageous and the highest efficiencies are to be gotten when there is "a smooth progression from individuals to commons." That is, the number of trunks in each subgroup of individuals, pairs, fours, eights and commons, should be very nearly equal, and in general "no grading should be used, if it can be avoided, in which the actual number of circuits required to the next rank exceeds half the maximum possible number.[9] An 18-group grading, for example, should not be used when more than 90 circuits are required. . . . These specific data are, of course, with reference only to the gradings of 10-contact switches."

Finally, a very interesting field of study has been opened up through the design of an artificial traffic machine by Messrs. E. A. Elliman and R. W. Fraser of the Standard Telephones and Cables, Ltd.[10] In the particular machine constructed two group gradings with all arrangements of individuals and commons up to hunts of 25 could be simulated. In the single examples given of straight and graded groups the results showing the load per trunk appear to be closely in accordance with their expected values. If a more flexible mechanism could be devised and tested to insure concordance with practice a most valuable contribution to the art of trunking would be made.

In the Mathematical Appendix I of this paper Mr. E. C. Molina sets forth the analytical theory for simple (single-stage) symmetrical graded multiples as originated and practiced (with certain modifications to be mentioned later) by the Bell System. The present method of estimating the probability of loss, knowing the arrangement of trunks and the average load submitted per subgroup, is the natural outgrowth of several preliminary formulas each of which was closely studied and compared with the actual conditions to be met in operation.

The four governing assumptions which need careful scrutiny in the final formula presented in this paper are:

1. The holding time of all calls is assumed to be constant.

2. A call not receiving immediate service is held in waiting for its normal holding time period, and if an available trunk becomes idle it will occupy it till this period is completed. This is usually referred to as the "lost calls held" assumption.

[8] G. F. O'Dell: "An Outline of the Trunking Aspect of Automatic Telephones." *The Journal of the Institution of Electrical Engineers*, vol. 65, February, 1927.

[9] By "next rank" is meant, for instance, second selectors following first selectors.

[10] Elliman and Fraser: "An Artificial Traffic Machine for Automatic Telephone Studies." *Electrical Communication*, October, 1928.

3. The distribution of the load submitted to each subgroup follows the Poisson Law and each subgroup carries the same average load as every other subgroup.

4. At no time shall a call be occupying a common trunk if an idle trunk exists in the group of individual trunks assigned to the subgroup of calling sources or switches from which the call under consideration originated. In other words, it is assumed that calls which seized idle common trunks because, at the time they originated, idle individual trunks were not available, shall be immediately transferred (by some fictitious redistributing apparatus) back to their individual trunks as soon as these become idle. This assumption will be referred to below as the assumption of "no-holes-in-the-multiple."

Whether these are admissible assumptions must be decided by a comparison of what actually results in practice with the formula which they give rise to. Their concordance will be discussed in a following section.

In order to put this graded formula in a usable form for engineering study curves or tables are needed showing how much load any given arrangement of trunks will be able to carry at any specified grade of service. Such charts have been constructed for the two more commonly used probabilities of loss, $P = .01$ and $P = .001$, comprehending all possible arrangements of trunks in simple symmetrical graded multiples having an access or assignment of 10, 20, 30 and 40 terminals and subgroups from two to seven in number. These are designated as Figs. 8 to 15, 8 to 11 corresponding to $P = .01$, and 12 to 15 to $P = .001$; the four at each probability cover the ranges of access or assignment, 10, 20, 30 and 40, respectively. These distinguishing parameters are noted in the upper right-hand corner of each chart. In order to simplify the necessary descriptive terms the number of trunks in each individual subgroup is called "x," the number of common trunks is called "y," and the number of subgroups, "g." Thus the access equals $x + y$, and the total number of trunks equals $gx + y$.

For the sake of compactness and brevity both the abscissa and ordinate scales of these charts are plotted in terms of ratios; the former gives $gx + y$ or total trunks in terms of the access, $x + y$, and the latter the per cent gain in efficiency (per cent increase in average load per trunk) over the efficiency of a simple straight multiple of $x + y$ trunks. A single one of the seven semi-circular curves on each figure then yields the load information for any number of trunks having a particular number of subgroups, a designated access, and a specified grade of service. The dotted curve on each figure is included to show

the "per cent gain over the efficiency of $x + y$ trunks" if the total $gx + y$ trunks are placed in a single complete-access group. A few problems will make clear the exact meaning and use of the curves.

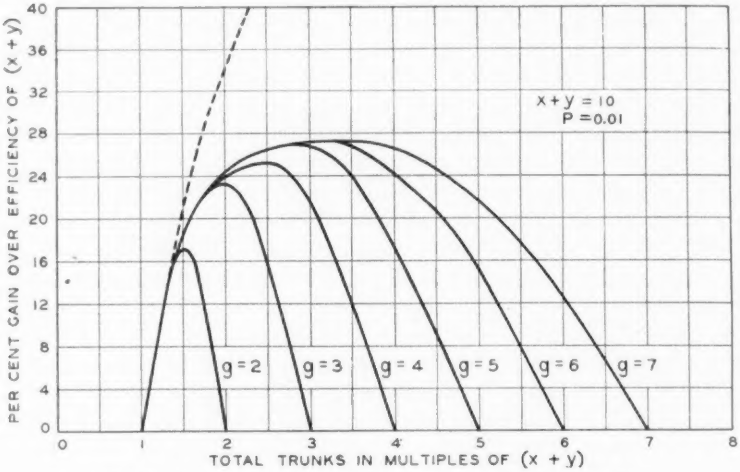


Fig. 8—Graded multiple efficiency. $x + y = 10$. $P = 0.01$.

Suppose we wish to know how much load may be submitted to each subgroup of a symmetrical graded multiple of 105 trunks having five commons out of an access of 30, such that, on the average, one call in

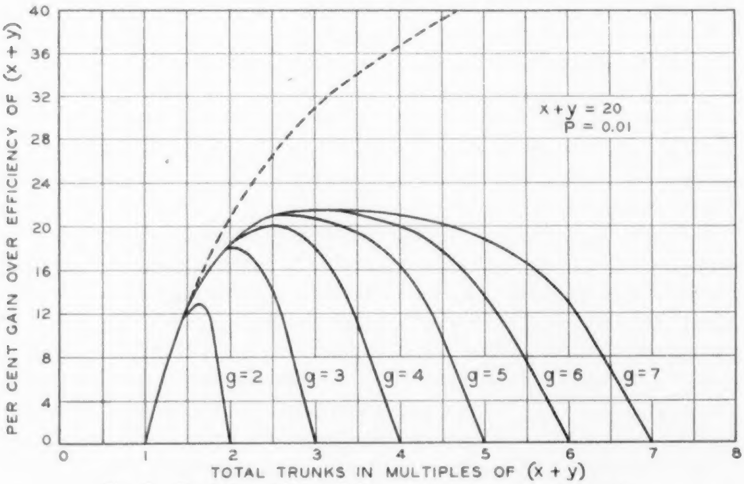


Fig. 9—Graded multiple efficiency. $x + y = 20$. $P = 0.01$.

one thousand will be lost. Evidently,

$$gx + y = 105, \quad x + y = 30, \quad x = 25, \quad y = 5, \quad P = .001;$$

from which, solving the first three equations simultaneously, we find g,

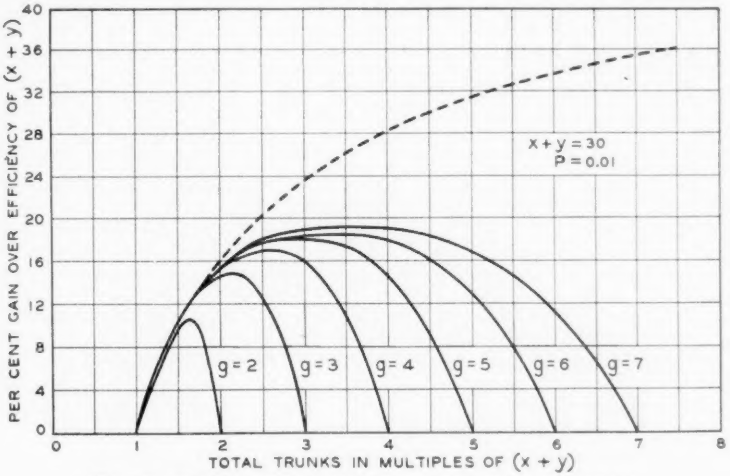


Fig. 10—Graded multiple efficiency. $x + y = 30$. $P = 0.01$.

the number of subgroups, equal to 4. To enter the appropriate chart (Fig. 14) we must determine the ratio $\frac{gx + y}{x + y}$, which equals $\frac{105}{30} = 3.5$. The $g = 4$ curve at this abscissa of 3.5 gives us the gain over the

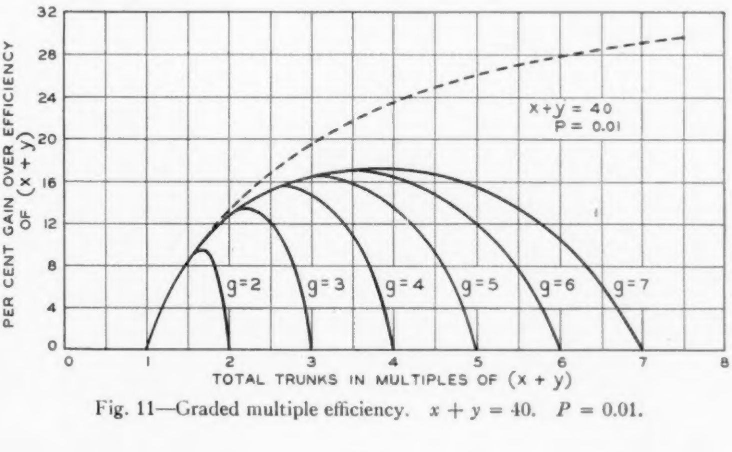


Fig. 11—Graded multiple efficiency. $x + y = 40$. $P = 0.01$.

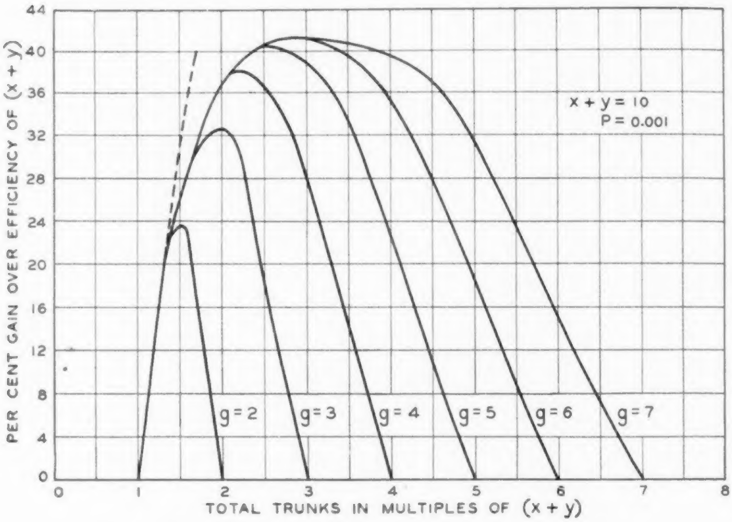


Fig. 12—Graded multiple efficiency. $x + y = 10$. $P = 0.001$.

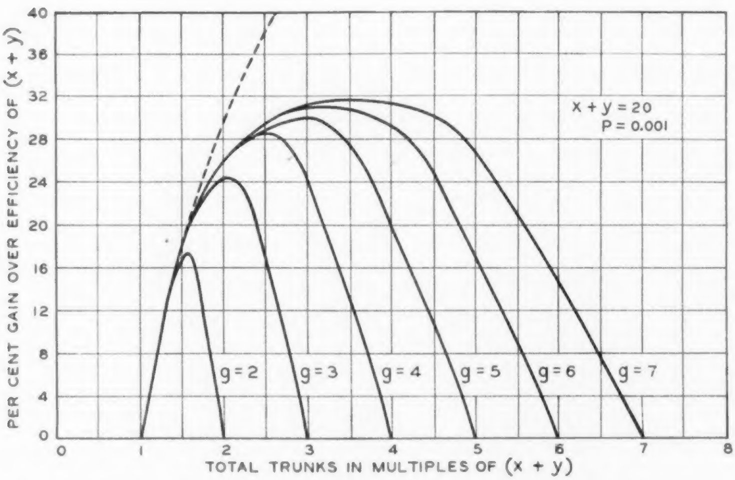


Fig. 13—Graded multiple efficiency. $x + y = 20$. $P = 0.001$.

efficiency of $x + y$ trunks as 13 per cent. This means that the average load per trunk in the graded multiple is 1.13 times that in a straight group of $x + y = 30$ trunks at the same probability of loss. From

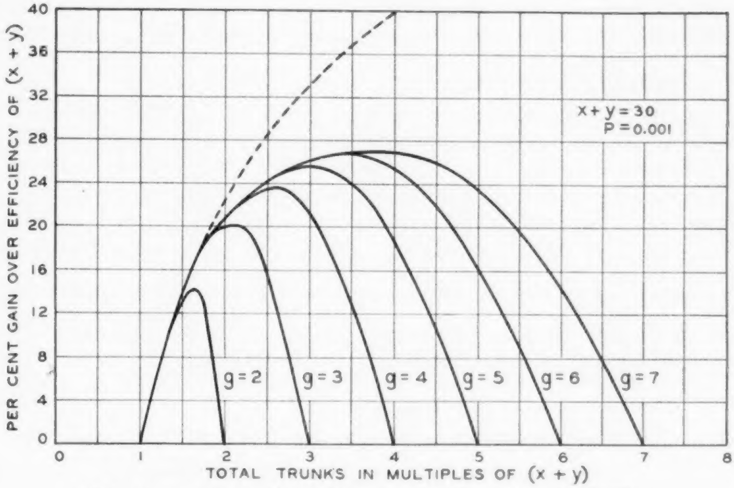


Fig. 14—Graded multiple efficiency. $x + y = 30$. $P = 0.001$.

Fig. 4 we read the average carried per trunk on a straight group of 30 as .529. Hence the total load which may be submitted [11] to the graded

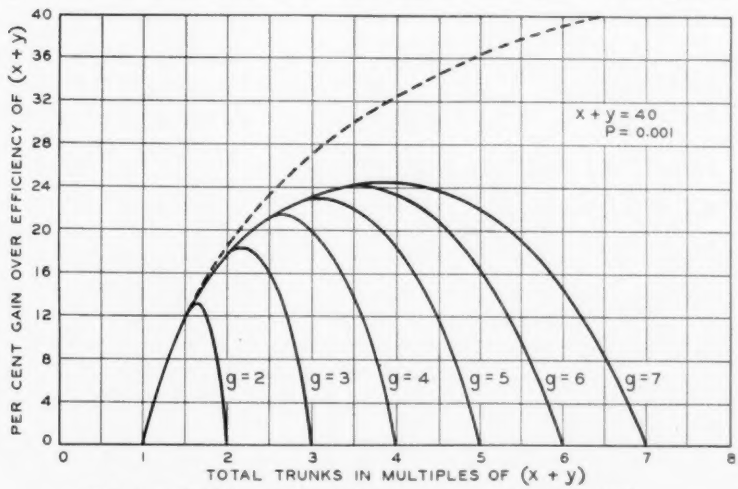


Fig. 15—Graded multiple efficiency. $x + y = 40$. $P = 0.001$.

[11] We may use the terms "average submitted" and "average carried" interchangeably here since at losses of $P = .01$ or less the difference is negligible.

multiple is

$$A = (1.13)(.529)(105) = 62.77.$$

To obtain the desired load per subgroup we need only divide by 4 giving $a = \frac{62.77}{4} = 15.69$. In terms of one-hundred-second-calls this average per subgroup then equals $36a = 565$ calls per hour.

As a second example we may set the question: Given 58 trunks to grade on an access of $x + y = 20$ with either three or six subgroups, which arrangement is to be preferred? We shall have to assume, since we have no additional information, that the decision rests merely upon the relative efficiencies of the two schemes. From the given data, $gx + y = 58$, $x + y = 20$, and $g = 3$ or $g = 6$. We read opposite $\frac{gx + y}{x + y} = \frac{58}{20} = 2.90$ for $g = 3$ and $g = 6$ on both the $P = .01$ and the $P = .001$ charts (since the probability was not specified) and construct the following table (Table II):

TABLE II

No. of Subgroups g	No. of Individual Trunks per Subgroup x	No. of Common Trunks y	$\frac{gx+y}{x+y}$	Per Cent Gain Over Efficiency of $(x+y)$	
				$P = .01$ Fig. 9	$P = .001$ Fig. 13
3	19	1	2.90	3.0	3.5
6	7.6	12.4	2.90	21.2	31.0

It is clear then, that as far as efficiency of arrangement goes, the six-subgroup plan is in the order of 20 per cent superior to the three-subgroup plan for ordinary grades of service. There is one difficulty here, however, and that is that in the symmetrical multiple of six subgroups the calculated number of individual trunks in each subgroup and the number of commons are not integers. We may get around this trouble by either unbalancing the grade slightly or changing the total number of trunks. For instance, we could use four subgroups of seven trunks, each pair of subgroups feeding into a single trunk common to them, before reaching the 12 through-commons into which the remaining two subgroups of eight trunks would work directly. The total of 58 trunks would then be disposed of at an efficiency gain probably not differing markedly from that estimated in our table above.

Secondly, we could have reduced the total trunks to 55, making thereby a symmetrical arrangement of $x = 7$ and $y = 13$. Had we

done this we should have been able, at $P = .01$, to carry a total load (reading the gain of 21.3 per cent at $\frac{gx + y}{x + y} = 2.75$ on Fig. 9) of $A_6 = (1.213)(.554)(55) = 36.96$. At the same time the load we could have carried on 58 trunks with three subgroups is only $A_3 = (1.03) \times (.554)(58) = 33.10$. Hence, we could reduce the total number of trunks by three in the six-subgroup case and still carry more load than if we were to use the three-subgroup arrangement and the original number of trunks. This decided advantage in favor of the larger number of subgroups is even more pronounced if the $P = .001$ comparison is made ($A_3 = 26.9$ vs. $A_6 = 32.2$).

The secret of the large gains in certain cases is easily found. After a short study of the curve charts it will readily be verified that, in general, the modal or maximum gain point on any curve comes very nearly at the midpoint of the range between $x + y$ and $g(x + y)$. This midpoint is reached by always setting $x = y$ or $x = \frac{1}{2}(x + y)$, that is, by making the individuals compose one-half of the access or assignment.

Comparisons of Theory and Practice

It is, of course, eminently desirable to know whether the formula just described for the probability of loss of any simple arrangement is consistent with the grades of service which it will actually render in practice. This we could ascertain only after a prolonged and careful study of typical graded groups already at work in the Bell System. Accordingly, a set of tests lasting over a period of six months, in the latter part of 1927, was made in Chicago by the Department of Operation and Engineering of the American Telephone and Telegraph Company in cooperation with the Illinois Bell Telephone Company.

The tests were performed on district multiples, believed to be representative, by connecting holding time recorders to the groups to indicate the load being carried by each trunk. Then through the use of overflow and peg count (number of calls) registers the proportion of calls being delayed (or lost) was readily found for any particular busy hour load.

In the First Division of these tests two groups of interoffice trunks, State to Dearborn and State to Wabash, were selected as typical cases of the kinds of fluctuating busy hour loads to be found in ordinary panel graded practice. No attempt was made here to regulate the load being submitted in any busy hour or to any subgroup since what was particularly desired was not what *would* happen *if* such and such conditions obtained, but rather what *does* happen under the fluctuating load conditions which *actually occur* in the busy hours from day to day.

TABLE III

Trunk Arrangements					Observational Data					Theory			
I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	XIII	XIV
										No. Trunks Required at the Observed Probability of Loss			
Total Trks. in Graded Group	Access or Assignment	No. of Sub-groups	No. of Individuals	No. of Commons	Average Load Submitted to Graded Group	Average Proportion of Calls Lost	Maximum Proportion Lost in Any One Hour	Per Cent of Total Busy Hours Having Overflows	No. of Busy Hours Included	Full Group Efficiency	Full Graded Gain	Half Graded Gain	No Gain
$gx+y$	$x+y$	g	x	y	A								
State to Dearborn													
28	20	5	2	18	19.50	.0573	.235	100.	13	26.9	27.4	28.1	28.9
32	20	5	3	17	18 97	.0106	.054	73.4	30	30.0	30.65	32.1	34.0
36	20	5	4	16	20.83	.0090	.075	69.1	81	32.8	33.2	35.2	38.0
40	20	5	5	15	20.77	.0010	.004	25.0	4	36.5	37.2	40.9	46.4
State to Wabash													
40	20	5	5	15	28.19	.0600	.169	100.	9	37.2	37.85	39.5	41.5
44	20	5	6	14	29.08	.0187	.041	100.	13	41.4	42.2	45.2	49.2
48	20	5	7	13	32.56	.0242	.110	100.	11	44.9	46.0	49.3	53.5
52	20	5	8	12	32.75	.0199	.111	84.8	66	45.5	46.9	50.5	55.0
56	20	5	9	11	35.25	.0136	.042	90.0	10	49.6	51.5	56.1	61.7

This will account then for the rather decided non-uniformity of the loads supplied to the two groups for various numbers of trunks. In Table III are recorded in the left-hand division the nine different trunking arrangements tested. All cases consisted of grades of five subgroups ($g = 5$) in a multiple having "x" individual trunks in each subgroup placed before "y" common trunks such that the access $(x + y)$ remained constant and equal to 20. The sizes of the resulting groups were then varied from 28 to 56 trunks.

In the central division of Table III, designated "Observational Data," are shown the various loads carried by these trunk arrangements with the corresponding proportions of calls lost during the period of each test. The supporting data of columns VIII, IX and X give one an idea as to the fluctuations which may be expected in the lost calls in a limited number of week-day busy hours.

The last division of Table III, "Theory," shows in its four columns the number of trunks that would theoretically be specified, on various bases of engineering, to carry the load actually observed in each run at a probability of loss equal to the observed proportion of calls lost. Column XI gives the number of trunks which would be required in each case could the trunks all be placed in a single straight group. Since this is the most efficient arrangement possible, a minimum of trunks need be supplied. At the other extreme we have in column XIV the number of trunks which would be required if each trunk operated at the efficiency of a group of 20 ($= x + y$) trunks. This could actually be realized, of course, only when the total number of trunks required was an exact multiple of 20. As shown, from 2 to 12 more trunks are required with this decreased efficiency than when a full group is being considered.

The two other columns, XII and XIII, show the number of trunks that would be needed in a graded multiple of five subgroups having an access of 20 trunks, upon two different assumptions. The column headed "Full Gain" is obtained from curves similar to Figs. 8 to 15, but appropriate to the observed probabilities of loss, to give the "per cent gain over the efficiency of $x + y$ trunks." Knowing the total load to be carried and the enhanced efficiency of $x + y$ trunks in each case the number of trunks required is readily determined. The "Half Gain" column is arrived at in precisely the same way with the exception that only one-half of the indicated "per cent gain over the efficiency of $x + y$ trunks" is utilized.

To facilitate the interpretation of these results they have been shown graphically in Fig. 16. Above each point on the abscissa at which a run with a known number of trunks was made is recorded

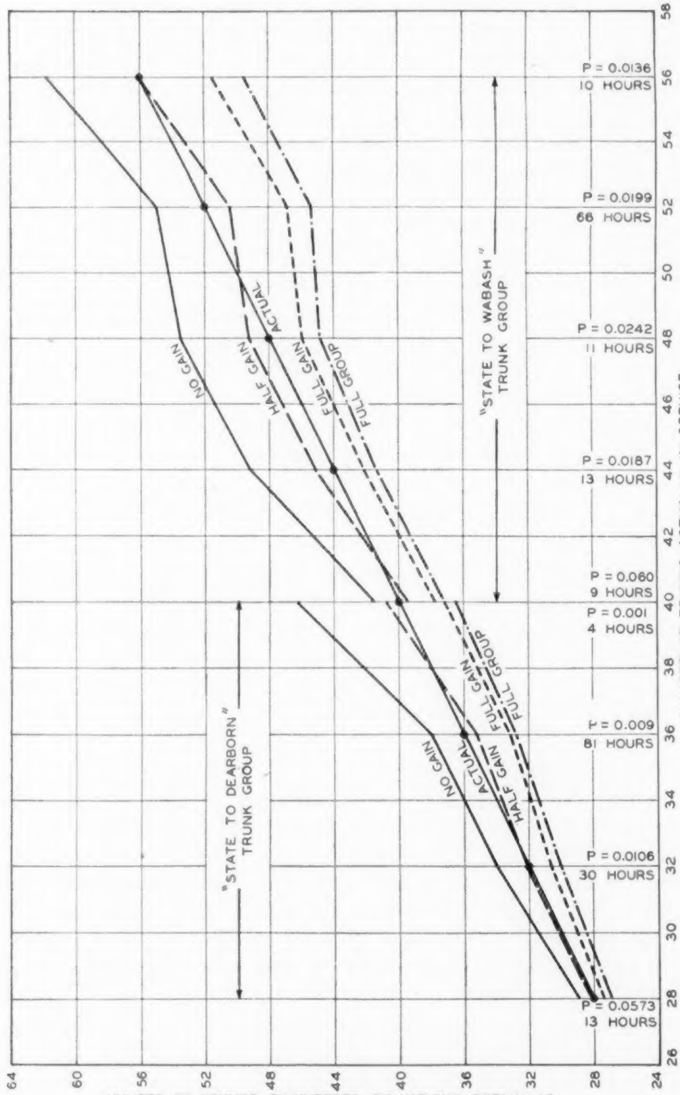


Fig. 16—Comparison of formulas with graded multiple busy hour tests, Chicago, July–December 1927.

the corresponding number of trunks which would be provided on each of the bases considered above. The actual number in service is also shown as a straight line through the shaded points for comparison with the various theoretical schedules. As noted on the figure the studies having 28 to 40 trunks were made on the State to Dearborn group and those having 40 to 56 trunks on the State to Wabash group.

As may readily be seen either from Table III or Figure 16 the "Half Gain" schedule coincides especially well with the observed data, while the other engineering plans fall consistently too high or too low on the scale. It may be remarked that the "Full Gain" values approach the limiting full group figures very closely and that it should prove of considerable interest to determine the cause of divergence between the field observations and these large theoretically possible graded loads.

Critical Inspection of Assumptions in Graded Theory

It has been noted that the number of "Full Gain" trunks specified is well below the number really required. This confirms the possible suspicion that not all of the four assumptions fundamental to the

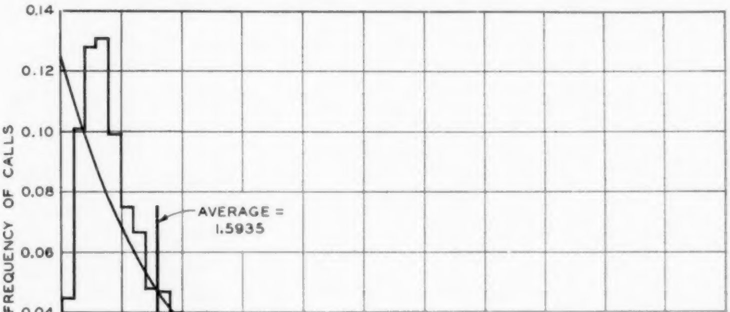


Fig. 17—Distribution of interoffice trunk holding times.

graded multiple formula presented in this paper are entirely satisfied. We shall therefore examine, in order, the accuracy of these assumptions.

We know indeed, without investigation, that calls are of widely differing lengths, and in addition are not "held" in the manner assumed. In Fig. 17 is shown a typical holding time distribution for

calls carried over the State to Dearborn group, with a suggested exponential fitting frequency curve superposed on it. From other studies carried out on non-graded groups we are led to believe that as far as the probability of loss is concerned it is almost independent of the change in holding time from a constant to an exponential form, the advantage, if any, favoring the varying case. Likewise the manner of "holding" delayed calls is of negligible importance as long as losses of .01 or .02 are not greatly exceeded.

The first part of the third assumption regarding the incoming calls being distributed according to the Poisson Law was not checked in these particular tests but a wide study of results under similar conditions readily leads us to believe that calls originating from a large number of independent sources will exhibit this form of frequency distribution.

The third assumption also necessitates a study of the variations among the loads submitted to the subgroups of a graded multiple. This brings us to the Second Division of the tests made in Chicago. At the same time the first division was in progress on the busy hours of each day for the cases of 36 trunks in service on the State to Dearborn group and 52 trunks on the State to Wabash group, all of the hours of the day from 9 till 5 were observed by one-half hour periods (to minimize the error due to trends) for the number of call-seconds on each trunk, the number of calls carried over the group and the number lost. Thus an extended range of load conditions was obtained for study. For estimating the effects of subgroup load variations with a given total submitted load, these short pieces of data were combined so that the half hours having an average load in trunk hours per hour within approximately one unit of range were thrown together. The various analyses were then made on these narrow total load classifications to discover, if possible, whether the observed subgroup variation when used in the theoretical formula would cause the latter's "Full Gain" probability of loss to approach more closely the observed losses.

First, the proportion of lost calls was determined for each of these approximate unit intervals of load. The results are shown graphically in Figs. 18 and 19 for State to Dearborn and State to Wabash, respectively. On these same figures have been superposed the theoretical curves for the losses to be gotten using "Full," "Half" and "No Gain" efficiencies in the graded formula described above. These theoretical computations have assumed that equal average loads are submitted to each subgroup at all times. The observed data indicate that the correct descriptive curves lie in both cases somewhere between those for the "Half" and "Full Gain" efficiency theories. This

seeming improvement of the observed data over its position in Fig. 16 is reasonable since here we have sorted out and combined only hours of like average loads while before all the busy hours, high and low, of a given period were included. It should be noted that the abscissa for the observations here is *load carried* while for theory it is *load submitted*. The comparison error is doubtless negligible for losses of,

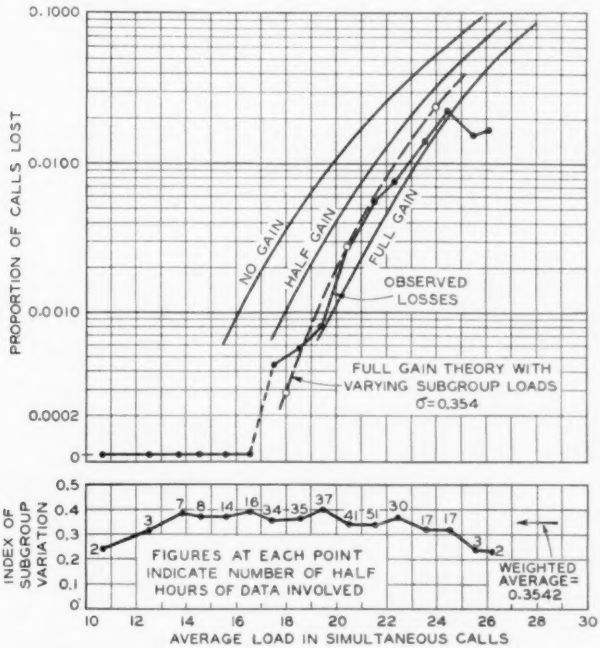


Fig. 18—Comparison of theoretical and observed graded losses, "State-Dearborn" group. $x = 4$, $y = 16$, $g = 5$.

say, $P = .03$ or less. The data beyond this figure are rather too meager for useful correction.[12]

Next the "Full Gain" formula was generalized by the author to comprehend the submission of different averages to each of the various subgroups. (See Appendix II.) Then, selecting typical loads, for instance $A = 18.00$ for the State to Dearborn tests and $A = 33.80$ for the State to Wabash tests, they were each divided up arbitrarily

[12] The advisability of correction here, were the data plentiful, may well be doubted since it would necessitate an assumption as to the manner in which calls were "held," a procedure especially precarious at high losses.

into five different magnitudes for use in this theoretical calculation. As a measure of the variation from equal loads submitted to the subgroups, the standard deviation of the estimated subgroup loads carried taken about their average value was used in each case. To estimate the load, "l," which would be carried by any subgroup to which the

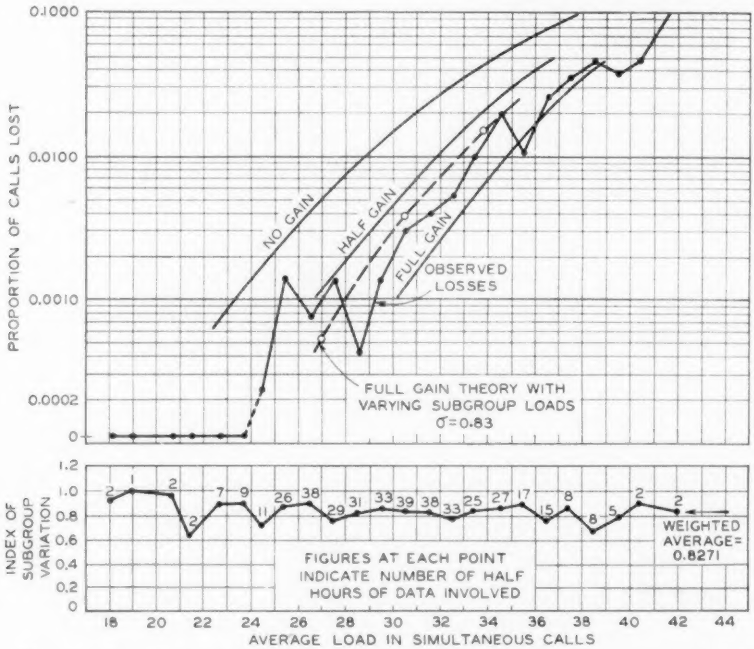


Fig. 19—Comparison of theoretical and observed graded losses, "State-Wabash" group. $x = 8$, $y = 12$, $g = 5$.

average, "a," is submitted, Erlang's formula, as set down earlier in this paper, was utilized in the following relationship:

$$l = a[1 - B(x, a)],$$

wherein, as before, $B(x, a)$ connotes the proportion of the submitted load which goes beyond the x individual trunks. The reason for working with the theoretical load carried by a subgroup rather than the load submitted to it will be clear when it is recalled that all the observed data available for comparison yielded the former values only. We shall now examine briefly the result of assuming subgroup variations

when the above typical total loads are substituted in the generalized graded formula.

TABLE IV

THE EFFECT OF SUBGROUP LOAD VARIATIONS ON THE GRADE OF SERVICE—36 TRUNKS

Total Load Submitted = A = 18, x = 4, y = 16, g = 5.

Case	Loads Submitted to Subgroups					Standard Deviation of Subgroup Loads Carried	Overall Probability of Loss
	a_1	a_2	a_3	a_4	a_5		
Theoretical No. 1....	3.60	3.60	3.60	3.60	3.60	0	.000189
Theoretical No. 2....	3.00	3.00	4.00	4.00	4.00	.18	.000223
Theoretical No. 3....	2.57	2.57	3.60	4.63	4.63	.35	.000289
Theoretical No. 4....	2.20	2.20	3.60	5.00	5.00	.48	.000531
Theoretical No. 5....	2.20	2.20	2.20	5.70	5.70	.58	.000957
Theoretical No. 6....	2.00	2.00	2.00	6.00	6.00	.67	.001493
Observed...........	—	—	—	—	—	.36	.00050

TABLE V

THE EFFECT OF SUBGROUP LOAD VARIATIONS ON THE GRADE OF SERVICE—52 TRUNKS

Total Load Submitted = A = 33.80, x = 8, y = 12, g = 5.

Case	Loads Submitted to Subgroups					Standard Deviation of Subgroup Loads Carried	Overall Probability of Loss
	a_1	a_2	a_3	a_4	a_5		
Theoretical No. 1....	6.76	6.76	6.76	6.76	6.76	0	.00619
Theoretical No. 2....	5.50	5.50	7.60	7.60	7.60	.45	.00925
Theoretical No. 3....	4.40	6.40	7.40	7.40	8.20	.69	.0116
Theoretical No. 4....	4.60	4.60	8.20	8.20	8.20	.89	.0181
Theoretical No. 5....	4.00	5.00	7.00	8.20	9.60	.99	.0214
Observed...........	—	—	—	—	—	.84	.013

In the last two columns of Tables IV and V are indicated the measures of subgroup load variation and the expected grades of service, respectively, for various theoretical unbalances studied on these two symmetrical graded multiples. Figs. 20 and 21 indicate the rapidity with which the overall probability of loss on the generalized "Full Gain" formula basis may be expected to rise with increases in the load unbalances in the subgroups. Reading off on the abscissa of Fig. 21 a measure of subgroup unbalance of .45, for example, indicates that for a total load of A = 33.80 being submitted to 52 trunks arranged in a grade of five subgroups and an access of 20, the correct probability of loss is not the "Full Gain" efficiency value of .00619 but rather the more conservative figure of about .0090.

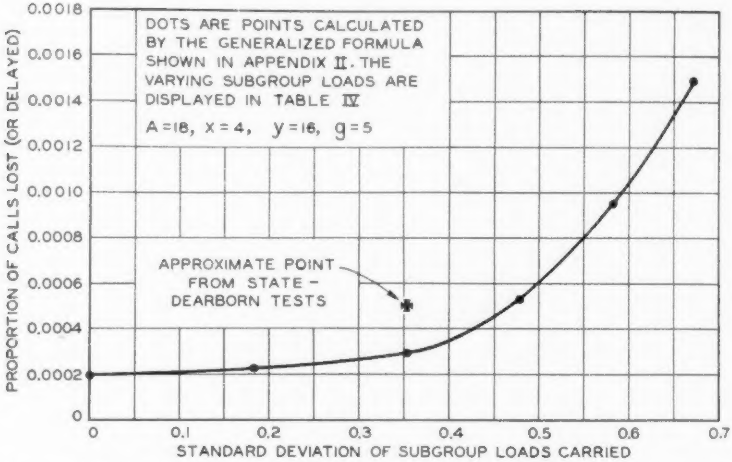


Fig.—20 Effect of subgroup load variation on probability of loss—36 graded trunks

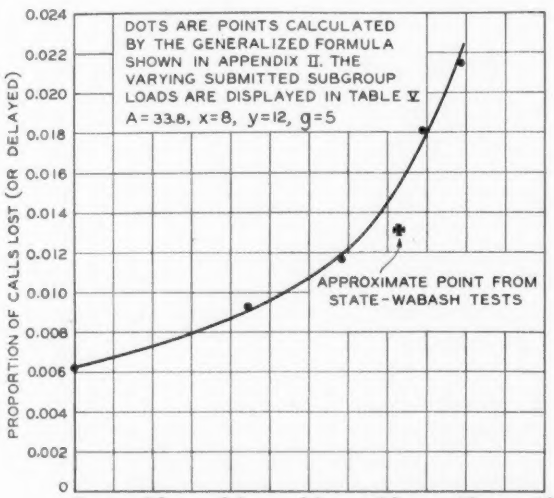


Fig. 21—Effect of subgroup load variation on probability of loss—52 graded trunks.

The corresponding subgroup variation obtaining throughout the Second Division of the tests was then determined and appears in the Figs. 18 and 19 as an auxiliary broken line at the bottom of the charts. Strangely enough, in both tests, the subgroup variation maintains an almost constant magnitude whatever the load. In fact, as mentioned in a later paragraph this criterion of variation seems rather to depend in practice upon the particular trunking arrangement being considered. Entering Fig. 20 with the observed average subgroup variation of .3542 for the 36 trunks in the State to Dearborn group, we should expect the proportion of calls lost at a total load of 18.00 submitted to be about .000289. Likewise, we should find that this same variation occurring in total loads of 20.5 and 24.0 gives us losses of .00278 and .0239, respectively. The dotted curve on Fig. 18 is drawn through these three points and represents the theoretical probability-load curve for this arrangement of trunks when the index of variation in loads from subgroup to subgroup is equal to the average observed. This schedule falls slightly above the observed losses over a considerable portion of the important range of loads although at the lower losses it seems quite likely to coincide fairly well with a curve drawn by eye through the data represented by the irregular line.

Similarly, in the case of the 52 trunks between State and Wabash, we may construct a schedule of losses based on a subgroup load variation having a measure equal to the observed value, .8271. Such a curve is shown dotted in Fig. 19 and as before indicates that upon taking into approximate consideration the variation in loads being submitted to the subgroups the resultant discrepancy between a curve fitted to the observed losses and the "Full Gain" theoretical grades of service is, in the main, slightly more than accounted for. In Figs. 20 and 21 the observed points fall one above and one below the theoretical schedules. These deviations do not appear to be of any significance except to illustrate the many chance elements which enter into telephone traffic problems.

A priori one might expect also that there would be some correlation between the proportion of lost calls and the subgroup variation for half hours in any given unit interval of total load. The number of calls lost in these tests, however, is so small that plotting by half hours the large natural fluctuation due to other causes seems to completely mask any such small effects which might be predicted.

To further study the manner and amount of this subgroup variation in carried loads, similar calculations were performed by half-hour units on the First Division of the tests (busy hours only) wherein the restriction of unit range of average was not present. The results shown by

the heavy dots in Fig. 22, as might be expected, give slightly higher values of variation for the non-restricted load values at $gx + y = 36$ and 52 trunks than for the restricted cases belonging to the second division of the tests. The remarkable point here, however, is that the phenomenon for the ranges studied exhibits practically a straight

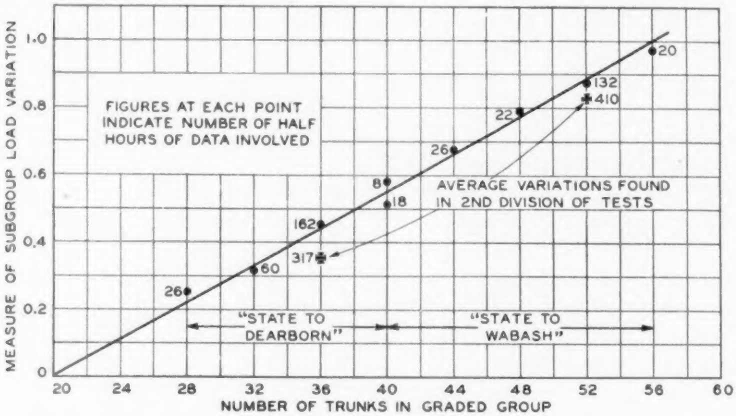


Fig. 22—Change in subgroup load variation with number of trunks in group.

line relationship between variation of subgroup loads and the number of trunks per subgroup, a difference of four in total trunks meaning an increase of one in each subgroup. An added variation for the larger number of trunks seems only natural, however, since as the subgroup size is increased part of the fluctuations previously borne by the commons is transferred to each subgroup itself. That this natural increase in subgroup variability does not affect the grade of service of the larger groups for busy hour measurements seems to be amply demonstrated by the consistency of the "Half Gain" formula in fitting the observed number of trunks in Fig. 16. We conclude, then, that a formula based on the third assumption (equality of subgroup loads), is considerably at variance with actual results; by modifying this assumption, to approximate loading differences, the graded loss formula appears to describe the observed losses quite satisfactorily.

Concerning the last assumption underlying the graded formula derived here, that of "no-holes-in-the-multiple," somewhat less is known. That the "holes" do exist is self-evident. It is suggested by these Chicago half-hourly observations, however, that under ordinary conditions the reaction of holes-in-the-multiple upon the grade

of service is negligibly small in comparison with the effect of subgroup load variations.

Practical Graded Multiple Engineering

In the Bell System, engineering of equipment and lines is done on the basis of the grade of service desired in the busy hours over a considerable period of time. The theoretical formula used, then, is the one giving, for those ranges in which satisfactory service is being rendered, an approximate relationship between the average total load carried and the proportion of lost calls expected over a number of busy hours. This, rather than some less conservative plan which would simulate the losses encountered only in a particular busy hour. The use of the "Half Gain" efficiency curves, therefore, is justified from a consideration of Fig. 16, which illustrates how closely that theoretical expression

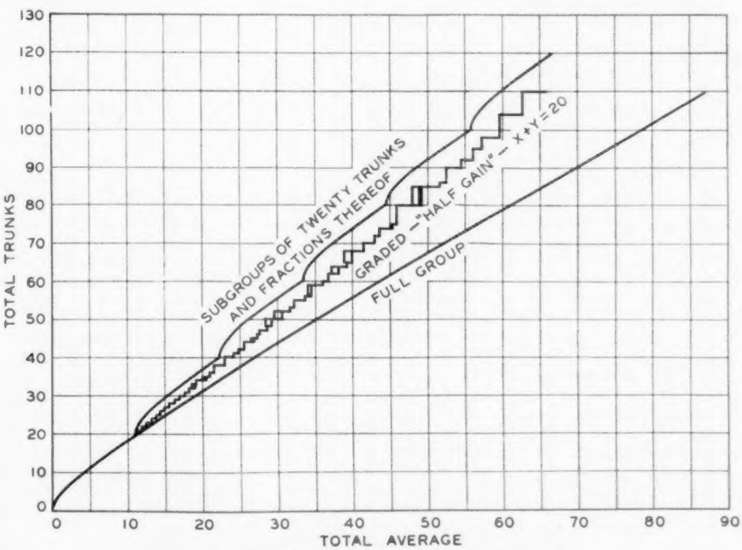


Fig. 23—Multiple capacity curves. $P = 0.01$.

approximates the actual conditions which maintain over a considerable range of trunk combinations and load values in two typical graded multiple installations.[13]

In practice, of course, the load to be carried rarely comes out exactly equal to that which a given symmetrical trunk arrangement

[13] Some five years ago it was appreciated that "Full Gain" was not likely to be attained. The value of "Half Gain" was then arbitrarily selected as the graded engineering basis in the Bell System until field studies of the efficiency could be completed. The merit of this estimate is attested by the material presented in this paper.

TABLE VI

GRADED MULTIPLE TRUNK CAPACITY TABLES

P = .01—Terminal Assignment = 20

Equivalent 100″ Calls	Number of Ind. Groups	Total Trunks	Ind. Trunks	Common Trunks	Equivalent 100″ Calls	Number of Ind. Groups	Total Trunks	Ind. Trunks	Common Trunks
423	2	21	1	19	954	3	44	12	8
448	2	22	2	18	961	4	44	8	12
448	3	22	1	19	961	5	44	6	14
470	2	23	3	17	961	7	44	4	16
470	4	23	1	19	987	6	45	5	15
492	2	24	4	16	992	3	46	13	7
492	3	24	2	18	1026	3	48	14	6
492	5	24	1	19	1030	4	47	9	11
515	2	25	5	15	1050	5	48	7	13
515	6	25	1	19	1065	3	50	15	5
540	2	26	6	14	1099	4	50	10	10
540	3	26	3	17	1099	6	50	6	14
540	4	26	2	18	1099	7	50	5	15
540	7	26	1	19	1140	5	52	8	12
562	2	27	7	13	1160	4	53	11	9
584	2	28	8	12	1215	6	55	7	13
584	3	28	4	16	1225	4	56	12	8
584	5	28	2	18	1236	5	56	9	11
609	2	29	9	11	1236	7	56	6	14
609	4	29	3	17	1300	4	59	13	7
634	2	30	10	10	1324	5	60	10	10
634	3	30	5	15	1324	6	60	8	12
634	6	30	2	18	1340	4	62	14	6
652	2	31	11	9	1370	7	62	7	13
670	2	32	12	8	1395	4	65	15	5
684	3	32	6	14	1410	5	64	11	9
684	4	32	4	16	1435	6	65	9	11
684	5	32	3	17	1494	5	68	12	8
684	7	32	2	18	1498	7	68	8	12
692	2	33	13	7	1544	6	70	10	10
717	2	34	14	6	1570	5	72	13	7
734	3	34	7	13	1627	7	74	9	11
740	2	35	15	5	1650	6	75	11	9
756	4	35	5	15	1650	5	76	14	6
756	6	35	3	17	1730	5	80	15	5
774	3	36	8	12	1757	6	80	12	8
774	5	36	4	16	1765	7	80	10	10
824	3	38	9	11	1857	6	85	13	7
824	4	38	6	14	1890	7	86	11	9
824	7	38	3	17	1958	6	90	14	6
872	3	40	10	10	2015	7	92	12	8
872	5	40	5	15	2055	6	95	15	5
872	6	40	4	16	2140	7	98	13	7
892	4	41	7	13	2258	7	104	14	6
918	3	42	11	9	2372	7	110	15	5

can carry at the allowable probability of loss. The next higher number of symmetrical trunks is then ordinarily specified. Thus, Fig. 23 shows the load-trunk curve for the "Half Gain" formula as a broken line representing the loads which may be submitted to the more im-

portant arrangements of $gx + y$ trunks for a terminal assignment of $x + y = 20$, g varying from two to seven subgroups, at a probability of loss of $P = .01$. This same figure portrays vividly the relative inefficiency of a straight subgrouped multiple compared with a like graded multiple, and again, the eminent superiority of the complete or full-group multiple over both of these. Table VI shows the same information as Fig. 23, recorded in the more familiar tabular form ready for engineering use.

Summary

We have sketched briefly some of the general principles underlying the furnishing of an adequate and economical telephone exchange service. One of the several practical means of interconnection is through the employment of special trunking arrangements of the type known as graded multiples. The common-sense theory of this plan has been discussed in some detail after which an approximate mathematical formula is presented.

A group of graded multiple tests run in Chicago in 1927 serves to indicate what modifications should be made in this theoretical trunking schedule before it is used for engineering purposes. A typical table of the loads which, on this basis, may properly be submitted to attain a specified grade of service over a wide variety of arrangements and numbers of trunks, is shown as an example of what appear as the most satisfactory graded multiple capacity figures for Bell System practice at the present time.

As a more detailed and accurate knowledge is acquired concerning the behavior of telephone traffic over increasingly complex and non-symmetrical graded trunking arrangements, some slight modification in the conclusions we have reached here may be expected. There is ample opportunity, then, for additional theoretical analysis of the graded multiple problem (as well as of many other multiple problems), an analysis that perhaps will overcome the limitations which have thus far been levied in order that a working result might be obtained.

Appendix I [14]

Mathematical Theory of the Simple Graded Multiple

The mathematical analysis given in this appendix is based on the following assumptions:

1. Constant holding time per call.
2. " Lost calls held."
3. The load submitted by each subgroup of selectors varies about its average value " a " in accordance with the Poisson Law

[14] Prepared by E. C. Molina.

$$\frac{a^x e^{-a}}{/x}$$

and is independent of the variations in any other subgroup.

4. "No-holes-in-the-multiple." [15]

The probability that a calling source fails to obtain an idle trunk immediately may be divided into two parts, corresponding to two essentially different sets of circumstances under which a call may be interfered with. Assume, to fix ideas, that the particular source under consideration belongs within subgroup No. 1 in Figure 7(A). Failure will occur if in addition to the call originating from this source,

1. At least $x + y$ calls originated by subgroup No. 1 occupy trunks (actually or potentially); or,
2. The number of calls placed on the trunks consists of $x + r$ originated from subgroup No. 1; at least $x + 1$ calls originated from each of s of the other subgroups; and, moreover, that said s other groups have, collectively, placed at least $sx + y - r$ calls. s is a number which may have any value from 1 to $(g - 1)$, inclusive.

This classification of the circumstances under which the call under consideration may be delayed gives, for the desired probability when "holes-in-the-multiple" are not possible, the equation

$$P = P_1 + P_2,$$

where

a. $P_1 = P(x + y, a)$

b. $P_2 = \sum_{r=0}^{y-1} \left(\frac{a^{x+r} e^{-a}}{/x + r} \right) F(g - 1, x, y - r)$

c. $P(x + y, a)$ is the Poisson expansion $\sum_{t=x+y}^{\infty} \left(\frac{a^t e^{-a}}{/t} \right)$

d. $F(g - 1, x, y - r)$

$$= \sum_{s=1}^{g-1} \binom{g-1}{s} [1 - P(x + 1, a)]^{g-1-s} \mathbf{S} \prod_{t=1}^{s} \left(\frac{a^{x+r_t} e^{-a}}{/x + r_t} \right),$$

where $\mathbf{S}$ indicates that in the product

$$\prod_{t=1}^{s} \left(\frac{a^{x+r_t} e^{-a}}{/x + r_t} \right), \; r_1, r_2, r_3 \cdots r_s$$

[15] Imagine that some method of transferring calls from common to individual trunks eliminates the possibility of "holes-in-the-multiple."

are to be given all values such that

$$r_t > 0, \qquad \sum_{t=1}^{s} r_t \not< (y - r).$$

For computing purposes, note that the function F satisfies the finite difference equation

$$\begin{aligned} F(g-1, x, y-r) = {} & P(x+y-r, a) \\ & + \sum_{R=0}^{y-r-1} \left(\frac{a^{x+}\, e^{-a}}{\underline{|x+R}} \right) F(g-2, x, y-r-R) \\ & + [1 - P(x, a)]F(g-2, x, y-r) \end{aligned}$$

This difference equation becomes obvious if one considers the change which takes place in the value of P_2 when the number of subgroups in a graded multiple is increased from $(g - 1)$ to g.

Appendix II

Mathematical Theory of Graded Multiple with Unequal Subgroup Loads

If we deal with a single stage (or simple) graded multiple having g subgroups of "x" individual trunks each, and "y" common trunks; and if to each subgroup, "m" for instance, is submitted a particular load, "a_m," in average simultaneous calls which are originated at random according to the Poisson Distribution Law requirements; and if these calls are moreover of a constant holding time obeying the "lost calls held" assumption; and if they are at all times so arranged on the graded trunks that "no-holes-in-the-multiple" exist; then the proportion of calls not obtaining immediate service over the multiple taken as a whole may be approximated by:

$$P = \frac{a_1 P_1 + a_2 P_2 + \cdots + a_g P_g}{a_1 + a_2 + \cdots + a_g} = \frac{\sum_{m=1}^{g} a_m P_m}{\sum_{m=1}^{g} a_m}$$

where

$$\begin{aligned} P_m = P(x+y, a_m) + \sum_{r=0}^{y-1} \frac{a_m{}^{x+r} e^{-a_m}}{\underline{|x+r}} F(g-1, a_1, a_2, \cdots, a_{m-1}, \\ a_{m+1}, \cdots, a_g, x, y-r), \end{aligned}$$

in which:

$$P(x+y, a_m) = \sum_{s=x+y}^{\infty} \frac{a_m{}^s e^{-a_m}}{\underline{|s}}$$

and

$$
\begin{aligned}
&F(g-1, a_1, a_2, \cdots, a_{m-1}, a_{m+1}, \cdots, a_g, x, y-r) \\
&\quad = P(x+y-r, a_1) + \sum_{R=0}^{y-r-1} \frac{a_1^{x+R} e^{-a_1}}{(x+R)!} F(g-2, a_2, a_3, \cdots, a_{m-1}, \\
&\qquad\qquad a_{m+1}, \cdots, a_g, x, y-r-R) \\
&\quad + [1 - P(x, a_1)] F(g-2, a_2, a_3, \cdots, a_{m-1}, a_{m+1}, \cdots, a_g, x, y-r).
\end{aligned}
$$

Moving-Coil Telephone Receivers and Microphones *

By E. C. WENTE and A. L. THURAS

A description is given of a moving-coil head receiver and a microphone designed particularly for high quality transmission. The instruments have a substantially uniform response from 40 to 10,000 c.p.s. This uniformity of response has been obtained, without sacrifice of sensitivity, by the use of light moving parts and the association of special types of acoustic networks with the diaphragm. In practical use the microphone has a sensitivity about 10 db higher than that of the Western Electric 394 Condenser Microphone.

MOVING-COIL loud speakers are now extensively used in high quality radio-receiving sets and in talking motion picture equipment. The chief advantages of the moving coil over the moving armature driving mechanism are the absence of a static force, constancy of force-factor and electrical impedance throughout a wide frequency range, and freedom from non-linear distortion over a wide amplitude range. Because of these advantages it seems obvious that the moving coil structure can also be used profitably in head receivers and microphones where high quality is of prime importance. It has therefore been adopted in the instruments to be described, although some of the principles here formulated can conceivably be applied also to instruments with moving armatures. This paper is concerned primarily with the general principles of design. The more practical phases of the commercial design and construction of the microphone are discussed in a paper by W. C. Jones and L. W. Giles.[1]

The moving system of a head receiver must, in general, satisfy distinctly different requirements from that of a microphone. In the actual use of the receiver a small enclosed cavity is formed between the ear and the diaphragm. If there is to be no distortion the pressure developed within this enclosure per unit of current in the receiving coil should be independent of frequency, constancy of impedance of the coil being assumed. The pressure depends not only upon the amplitude of vibration of the diaphragm, but also upon the acoustic impedance of the cavity formed by the ear and the receiver. This impedance is such, if the cavity is entirely enclosed, that at low frequencies the pressures will be very nearly proportional to the displacement of the diaphragm. At higher frequencies it is of uncertain value and varies from ear to ear, but it appears, from unpublished

* Jour. Acous. Soc. Amer., July, 1931.

[1] "Moving Coil Microphone for High Quality Sound Reproduction." Presented at May 1931 meeting of Soc. of Motion Picture Engineers, Hollywood, California.

data obtained by L. J. Sivian on a large number of ears, that constant amplitude of motion of the diaphragm per unit current throughout the frequency range is on the average the best condition to strive for in the design of a high quality receiver. We shall therefore assume that at any frequency the amplitude of motion of the diaphragm per unit current is a correct measure of the response of the receiver. It will be assumed also that the impedance of the cavity is without effect on the displacement of the diaphragm. For the receivers to be considered this assumption introduces but little error, although the effect is not negligible in general.

The voltage generated by a moving coil in a magnetic field is proportional to the velocity; therefore, the diaphragm of a uniformly sensitive microphone with a rigidly attached coil should have, at all frequencies, the same velocity per unit of pressure in the actuating sound wave. Expressed in another way, if the diaphragm has a constant effective area, the mechanical impedance (force per unit velocity) of a microphone diaphragm should be the same at all frequencies, whereas that of the receiver should be inversely proportional to the frequency. The receiver and the microphone to be described are quite similar in design and construction, but their dynamical constants differ so as to approach these conditions of impedance.

If a receiver or microphone is constructed with a diaphragm having a single degree of freedom, the operating conditions of the diaphragm can be represented by the circuit diagram shown in Fig. 1, where m_0

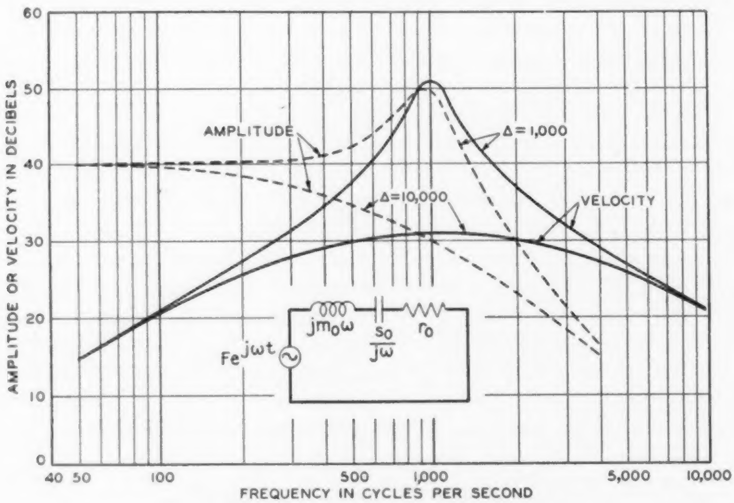


Fig. 1—Response of a simple resonant system.

is the effective mass, s_0 the stiffness, r_0, the mechanical resistance of the diaphragm, and $Fe^{j\omega t}$ the alternating force acting upon the diaphragm. The absolute value of the velocity of the diaphragm is given by

$$v = \frac{F}{m\left[4\Delta^2 + \left(\frac{\omega^2 - \omega_0^2}{\omega}\right)^2\right]^{1/2}}$$

and the amplitude by v/ω where $\Delta = \frac{r_0}{2m_0}$, the damping constant, and $\omega_0 = \sqrt{\frac{s_0}{m_0}} = 2\Pi \times$ resonant frequency. The velocities and amplitudes for a constant force and for two different values of Δ, calculated from these expressions, are graphically represented in Fig. 1. Both the amplitude and velocity curves show wide variations in response with frequency. They indicate that for small variations in amplitude the resonant frequency must be near the upper limit of the frequencies to be transmitted and for small variations in velocity the damping constant must be high. But instruments designed on this basis would be relatively insensitive even if such conditions could be met readily in their construction.

In the design of electrical networks for the transmission of wide frequency bands the end is attained by the combination of more than one resonant circuit. We can advantageously resort to a similar expedient in a mechanical system by the use of a structure more complicated than one having a single degree of freedom. The diaphragm may be coupled to another mechanical or acoustical network of the proper type so as to give us the desired uniformity of response. The circuit diagram of one such mechanical network is shown in Fig. 2,

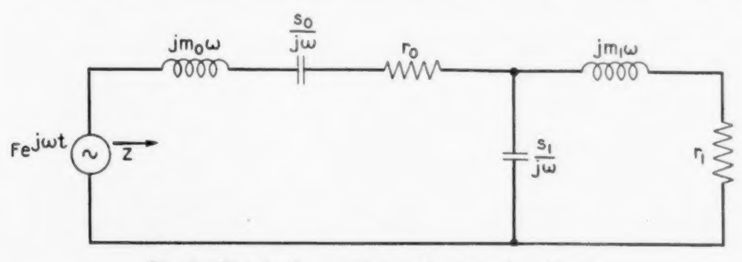


Fig. 2—Circuit diagram for receiver or microphone.

where s_1 is the stiffness, m_1 the mass, and r_1 the resistance of the elements of the coupled network. The construction of a mechanical system represented by this diagram is brought out in detail in the

discussion of the mechanical design of the instruments which is to follow. The actual values of the constants are to be so chosen, if possible, that the mechanical impedance, z, of the whole network is constant with frequency in the case of the microphone and inversely proportional to the frequency for the receiver. The absolute value of this impedance, z, is $\sqrt{r^2 + x^2}$ where

$$\left.\begin{aligned} r &= \frac{s_1^2 r_1}{r_1^2\omega^2 + m_1^2(\omega_1^2 - \omega^2)^2} + r_0, \\ x &= \frac{s_1\omega[m_1^2(\omega_1^2 - \omega^2) - r_1^2]}{r_1^2\omega^2 + m_1^2(\omega_1^2 - \omega^2)^2} + m_0\omega - \frac{s_0}{\omega}, \\ \omega_1^2 &= \frac{s}{m_1}\,. \end{aligned}\right\} \tag{1}$$

The Electrodynamic Receiver

If the mechanical system of the receiver can be represented by the circuit diagram shown in Fig. 2, then, as the amplitude per unit force is a measure of the receiver response, we may calculate the product of frequency and impedance and so get a response-frequency characteristic for any specified set of values of the constants. Such characteristics are graphically shown in Fig. 3 for several sets of values. Curves

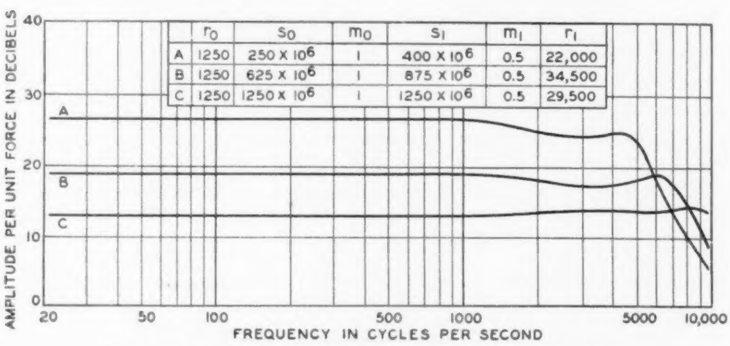


Fig. 3—Theoretical response curves of moving coil receiver.

of identical character but of different level would, of course, be obtained if the magnitude of each of the corresponding impedance elements were changed in the same proportion. It is seen from these curves that, theoretically at least, it is possible to obtain a uniform response over a wide frequency range. Curve C, for example, shows a variation of less than 1.5 db for frequencies up to 10,000 c.p.s. As might be expected, the wider the frequency range of uniform response

the lower the sensitivity. In fact, it can be shown from equations (1) that, if the scale of frequencies is changed by a factor, k, the relative values of the ordinates will be unchanged provided r_1 and r_0 are multiplied by k, and s_1 and s_0, by k^2, but that the amplitude per unit of force at corresponding points on the curve will be changed by a factor equal to $1/k^2$. A receiver transmitting up to 10,000 c.p.s. will thus be 12 db less efficient than one transmitting equally well up to only 5000 c.p.s., the same mass and size of diaphragm being assumed.

Construction of the Receiver

The general construction of a receiver embodying the above principles is shown in Fig. 4. The central portion of the diaphragm is

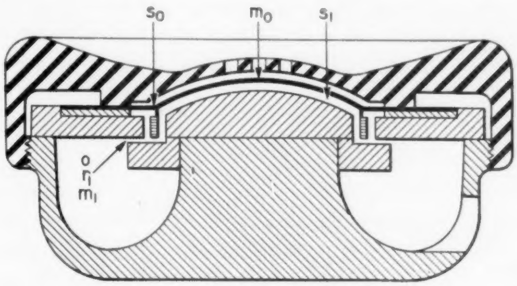


Fig. 4—Moving coil head receiver.

drawn into the form of a spherical dome to increase its rigidity. The receiving coil is of the self-supporting ribbon type, the construction of which has been described previously.[2] It is rigidly attached to the base of the domed portion of the diaphragm. The radial magnetic field is derived from a permanent magnet. The mass of the diaphragm plus that of the coil corresponds to m_0 in Fig. 2, the stiffness of the diaphragm to s_0 and the mechanical resistance to r_0.

A small volume of air is completely enclosed between the diaphragm and pole-pieces save for a narrow slit at O. The acoustic resistance [3] of a slit of this character is equal to $\dfrac{12\mu l}{d^3 w}$ and the reactance, $j\dfrac{6}{5}\dfrac{\rho l}{wd}\omega$, where μ is the viscosity of air, l the radial length, d the width, w the annular length of the slit and ρ the density of air. If the air in the chamber were incompressible a mechanical resistance and reactance would be imposed on the diaphragm by virtue of the air flow through the slit, their respective values would be equal to the acoustic resistance and

[2] *Bell System Technical Journal*, Vol. VII, p. 144, 1928.
[3] Lamb "Hydrodynamics," 4th ed., p. 577.

the acoustic reactance of the slit multiplied by the square of the effective area of the diaphragm. These quantities are represented by r_1 and $j\,m_1\omega$ in Fig. 2. If the slit, O, were closed the stiffness imposed by the air chamber on the diaphragm would be equal to $\frac{\gamma A^2 10^6}{V}$, where A is the effective area of the diaphragm, V the volume of air in the enclosure and γ the ratio of specific heats of air. This is the stiffness represented by s_1 in Fig. 2.

In adjusting the width of the slit to the desired value its resistance was measured experimentally. For this purpose a steady stream of air at low velocity was passed in series through the slit and a capillary tube. The pressure drop through the tube and that through the resistance was then measured with a manometer. The ratio of these values is under this condition equal to the ratio of the resistance of the tube to that of the slit. The resistance of the tube had previously been determined as a function of the pressure difference between its two ends when air was passed through it at a known steady rate. The apparatus is diagrammatically shown in Fig. 5.

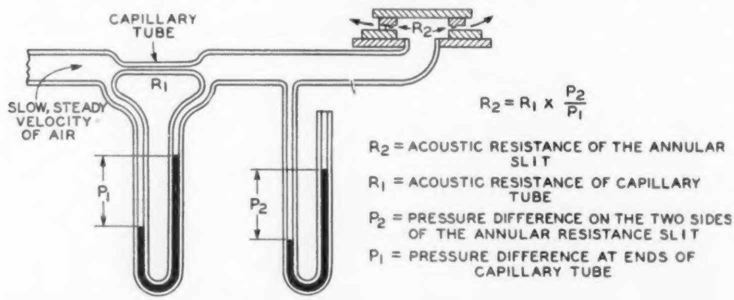


Fig. 5—Method used to measure acoustic resistance.

The response-frequency characteristic of the receiver was determined experimentally. For these measurements it was placed over a calibrated condenser microphone so as to form a 15 c.c. enclosure between the receiver and the microphone diaphragms. This space was filled with hydrogen to avoid acoustic resonance at the higher frequencies. While current from a vacuum tube oscillator was passed through the receiver coil, the voltage generated by the microphone, as well as the receiver current, was measured. From these values, the calibration curve of the microphone and the volume of the enclosure, the amplitude of the receiver diaphragm per unit current is readily determined. Values so obtained, expressed in db, are plotted in Fig. 6. In the

same figure are given values of the response as determined by computation of the mechanical impedance from the constants of the receiver. The ordinates were so adjusted arbitrarily as to bring the computed and observed values into coincidence at the lower frequencies. There is a general agreement between the computed and ob-

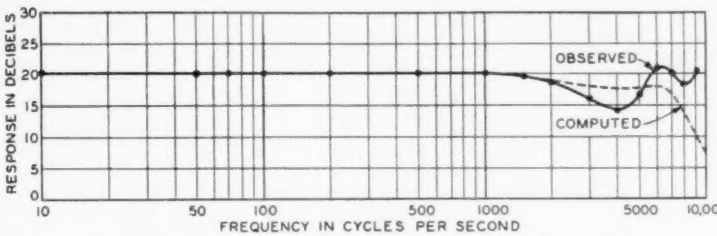


Fig. 6—Response of moving coil receiver.

served curves, yet the variations are larger than can be accounted for on the basis of experimental errors. It is probable that the quantities used in the calculations are not strictly constant up to the higher frequencies, where the diameter of the diaphragm becomes comparable to the wave-length of sound. However, except for a depression in the neighborhood of 4000 c.p.s., the measured is better than the computed characteristic.

A receiver of this general character was supplied for the Master Reference Systems for Telephone Transmission [4] in Europe and in America where it has been in service since 1928.

The Moving Coil Microphone

It has been pointed out that in an electrodynamic microphone of high quality the diaphragm with a rigidly attached coil should have the same velocity per unit of force throughout the frequency range. If the dynamical system of the microphone is represented by the mechanical circuit of Fig. 2, this condition requires that the constants of the various elements of this circuit be so chosen that the magnitude of the impedance, z, is the same at all frequencies. It is evident that these values will differ materially from those of the high quality receiver.

In Fig. 7 the impedance expressed in db as determined by equations (1) is shown as a function of frequency for several sets of values of the constants of the impedance elements. They show how, by the proper choice of these values, a uniform response may be obtained over a

[4] "Master Reference System for Telephone Transmission," by W. H. Martin and C. H. G. Gray, *Bell Sys. Tech. Jour.*, July, 1927.

wide frequency range. Curve C, for instance, shows a variation of less than 1.5 db from 200 to 10,000 c.p.s. It may be shown from equations (1) that if the scale of frequencies is changed by a factor k the form of the response curve will remain unchanged, provided r_1 and r_0 are multiplied by k, and s_1 and s_0, by k^2; but the absolute value of the velocity per unit force will be changed by the factor $1/k$ at all points on the curve. Thus, under these conditions, if the last value of the abscissæ in Fig. 7 is designated as 5000 instead of 10,000 c.p.s.,

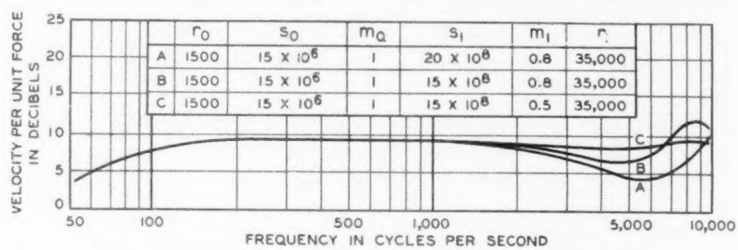


Fig. 7—Theoretical response curves of moving coil microphone.

then $k = 0.5$ and the curves will remain unchanged in form but the ordinates will be raised 6 db. The form of any of the curves of Fig. 7 will, of course, not be changed if all the corresponding constants are changed proportionally, although the absolute value of the velocity per unit of force will vary inversely with the magnitude of these constants.

At zero frequency the velocity of the diaphragm per unit of force is necessarily zero. In passing to the lower frequencies a point is therefore finally reached where the response decreases appreciably. This point depends primarily upon the stiffness, s_0, of the diaphragm. A method for overcoming this loss in sensitivity at low frequencies will be discussed later.

Construction of the Microphone

A microphone was constructed very similar in design to that of the receiver just described, but with the cap omitted in order to expose the diaphragm to the action of sound waves. The dimensions of the various elements were changed so that the impedance of the diaphragm with its associated network should have a substantially constant value throughout a wide frequency range. The response as computed is shown in Fig. 8.

The moving coil microphone was calibrated experimentally by comparison with a calibrated condenser microphone. For this comparison each transmitter was mounted with its face outward in an opening in the end

wall of a cylindrical drum 30 cm. in diameter and 7 cm. deep. The two openings were spaced 180° with respect to the axis of the drum and on radii of 7.5 cm. Cracks between the microphones and the wall were carefully sealed. The wall thus formed a baffle of the same general character for each microphone. The drum was mounted on a shaft passing through its axis, about which it was rotated at a speed

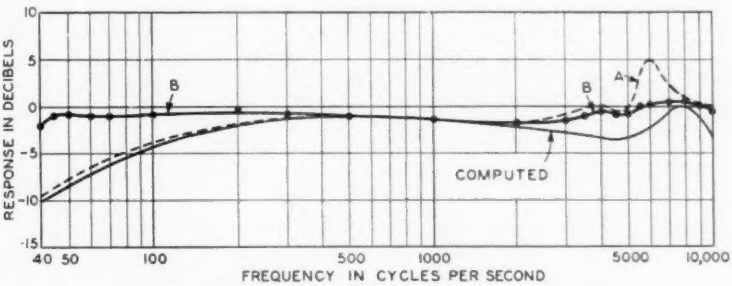


Fig. 8—Response of moving coil microphone.

of 100 r.p.m. Slip rings were provided for making electrical connections to the microphones. The drum was placed in a sound field set up by a moving coil loud speaker supplied with current from a vacuum tube oscillator. The voltage generated by each microphone was then measured with an amplifier and thermo-galvanometer. With this arrangement each microphone passed through practically the same sound field. By virtue of the symmetrical character of the drum its rotation has very little influence on any standing wave patterns in the room. A check on the reliability of the measurements was the fact that, if the position of the loud speaker was changed very little difference was observed in the ratio of the voltages generated even at the higher frequencies. Likewise, no change was observed when the electrodynamic microphone was moved a small distance axially in its mounting. The condenser microphone used in these tests had been calibrated by means of a thermophone, but a correction was made for the resonance due to the cavity over the face of the diaphragm, which is not measured in the thermophone calibration. The response of the microphone as determined in this way is shown by the curve A in Fig. 8.

The disagreement between the observed and computed values at the higher frequencies is believed to be due to resonance oscillations within the air-chamber beneath the diaphragm. In order to reduce the magnitude of these oscillations the chamber was connected through a narrow slit r_3 (Fig. 9) to a small cavity formed within the central

pole-piece. With this change in construction, the microphone was again calibrated. The results obtained in this case at the higher frequencies are shown by curve *B* of Fig. 8.

It is seen that the response of the microphone is quite uniform over a wide frequency range, but that it decreases at the lower frequencies. This decrease can be avoided by a reduction in the stiffness, s_0, but this

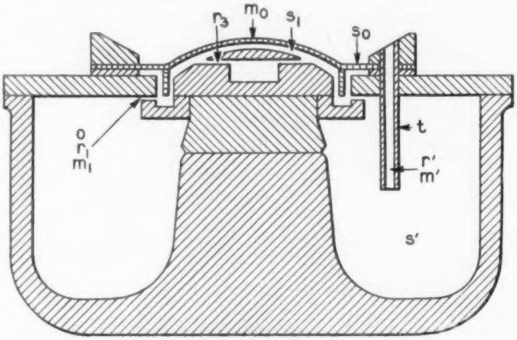


Fig. 9—Moving coil microphone.

expedient has the practical disadvantage that it makes the transmitter more delicate and increases its susceptibility to mechanical vibrations. The response at these frequencies can be increased more profitably by a simple modification which increases the force on the diaphragm under the action of sound waves. If the air-space enclosed by the magnet on the rear of the diaphragm is connected with the outside air through a tube, then, under the action of sound, a pressure will be developed within this space through the tube, differing in magnitude and phase from that of the sound outside. This pressure acts on the rear of the diaphragm. Under certain circumstances the total force on the diaphragm will be increased by virtue of this pressure.

The microphone shown in Fig. 9 is provided with a tube for performing this function. The acoustic impedance of a tube may be calculated from the formula [3]

$$Z = \frac{-\mu k^2 l}{\Pi r^2}\left[\frac{1}{1 + \dfrac{2\,J_0(kr)}{k\,J_0(kr)}}\right], \tag{2}$$

in which $k = \sqrt{j\mu/\rho\omega}$, l is the length and r the radius of the tube, μ the viscosity and ρ the density of air. At low frequencies, Z may be

[3] I. B. Crandall, "Theory of Vibrating Systems and Sound," p. 237.

represented by a resistance in series with a mass reactance, and the whole dynamical system of the microphone by the circuit diagram of Fig. 10, in which $Fe^{j\omega t}$ is the pressure in the sound wave multiplied

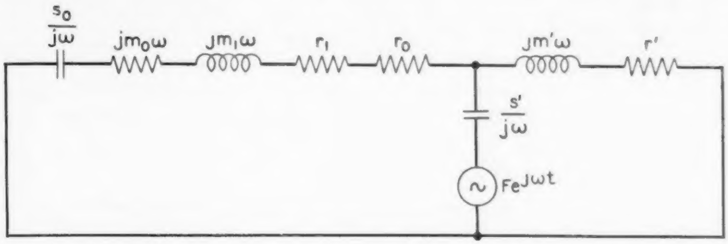


Fig. 10—Simplified circuit diagram of moving coil microphone at low frequencies.

by the effective area of the diaphragm; s', the stiffness imposed upon the diaphragm by the air enclosed within the magnet, if the tube were closed; r' and m' are the acoustic resistance and mass respectively of the tube multiplied by the square of the area of the diaphragm. The other symbols of Fig. 10 have the same significations as before.

Substituting numerical values for the various impedances of the circuit shown in Fig. 9, and solving this circuit for the velocity of the diaphragm per unit of force, we obtained the low-frequency values given by the curve B of Fig. 8. The circles give the corresponding values obtained experimentally. The agreement between these values and those computed is within the experimental errors with which the constants of the microphone were determined. The addition of this acoustic network has increased the response at the low frequencies so that there is no loss in sensitivity down to a frequency of 45 c.p.s., even with a diaphragm of comparatively high stiffness.

The absolute sensitivity of this microphone is approximately 9.5×10^{-5} volts per bar. However, in practical operation a transformer is used between the microphone and the vacuum tube of the initial stage of the amplifier. The transformer that has been used for this purpose has a voltage ratio of 100 with a variation of less than 2 db between 45 and 10,000 c.p.s. Under this condition the voltage delivered to the vacuum tube is 9.5 millivolts per bar. This value compares with approximately 3 millivolts per bar for the Western Electric Company 394 Condenser Microphone, which was designed for maximum efficiency for frequencies up to 7,000 c.p.s. The electrodynamic microphone thus has a sensitivity about 10 db higher, and covers a wider frequency range.

The condenser microphone commonly used has a cavity in front of

the diaphragm. Acoustical resonance in this cavity increases the pressure on the diaphragm, which in the case of the W. E. Co.'s 394 Transmitter may, under certain circumstances, amount to 5 db at a frequency of 3500 c.p.s. The microphone here described is believed to be relatively free from this effect, as the cavity in front of the transmitter is conical and quite shallow. The diaphragm is also smaller, so that the response is uniform over a wider angle of sound incidence.

This microphone has important practical advantages over the condenser microphone in that the amplifier may be at some distance from the microphone without loss in efficiency and in that no polarizing voltage is required. The sensitivity of this microphone is about 10 db higher. It is therefore better adapted for use in cases where the source of sound is at some distance from the microphone, since, with the smaller amplification required, mechanical and electrical disturbances, and amplifier noises in general, may be kept at a relatively lower level.

Some Developments in Common Frequency Broadcasting *

By G. D. GILLETT

This paper describes the results of the simultaneous operation of radio stations WHO and WOC broadcasting the same program on a common frequency using independent crystal controlled oscillators. These stations had previously been compelled to share time on 1000 kc and each is now able to render full time service.

The exceptional stability of the crystal controlled oscillators used at each station is described. Since even these oscillators require occasional readjustment to maintain them in isochronism, a monitoring receiver was established midway between the stations and the resultant program is sent back by wire line to WOC to provide an indication for readjusting its frequency to exact isochronism with WHO. An audio oscillator used to modulate the carriers in the monitoring receiver provides a tone independent of the program for the guidance of the operator. Curves are presented showing the quality impairment caused by different degrees of isochronism and signal strength ratios.

The improvement in distant reception with simultaneous operation is reported and an explanation given. The impaired reception in the area midway between the stations and outside their normal service range is shown to be a function of the degree of modulation of each transmitter, of the field strength ratio and of the audio phase angle and independent of the carrier phase at the transmitters. It is pointed out that reception equal to that from either station alone may still be obtained in this area by the use of a simple directive antenna.

The marked increase in the service rendered by these stations through simultaneous operation is indicative of the improved service that can be rendered to urban areas by common frequency broadcasting. Although it is probable that the high powered station on a cleared channel will remain the best means of affording a high-grade service to a metropolitan area while also rendering an acceptable service to large rural areas, common frequency broadcasting now appears to offer definite means by which to provide an improved coverage to a number of noncontiguous communities.

THE development of chain broadcasting and the congestion in the broadcast frequency range has naturally led to a consideration of the possibilities of operating a group of stations on a single frequency.[1,2] The possible usefulness of such a system has resulted in a number of attempts to secure the additional coverage offered by the simultaneous operation of two or more stations broadcasting the same program on a common frequency. This problem has been attacked in two different ways.

* Presented at Sixth Annual Convention of I. R. E., June 4–6, 1931. Published in Proc. I. R. E., 19, 1347–1369; August, 1931.

[1] DeLoss K. Martin, Glenn D. Gillett, and Isabel S. Bemis, "Some Possibilities and Limitations in Common Frequency Broadcasting," *Proc. I. R. E.*, **15**, 213–223; March, 1927.

[2] Charles B. Aiken, "The Detection of Two Modulated Waves which Differ Slightly in Carrier Frequency," Proc. I. R. E., January, 1931; B. S. T. J., January, 1931.

In one case, a control frequency has been transmitted either by wire line or radio to each station and a frequency multiplier used to develop directly the carrier frequency which was to be transmitted from the station. This method has met with some success both here and abroad. It was used in this country for the commercial operation of WBZ-WBZA [3] and in Germany the Postal authorities have operated several stations experimentally with equipment developed by the Telefunken G.m.b.H. and the C. Lorenz A.G.[4,5] Both the WBZ-WBZA and the Telefunken systems used a high control frequency which was particularly suitable for transmission over open wire lines while the Lorenz system used a lower control frequency which was suitable for transmission over cable circuits as well. Three stations located at Berlin, Stettin, and Madgeburg, respectively, are now in commercial operation on a common frequency using control equipment manufactured by the Lorenz firm.[6] In Sweden the postal authorities have developed a similar system of frequency control capable of using either a high or low standard frequency interchangeably. This system was used in placing the broadcast stations at Malmo and Halsingborg in commercial operation on a common frequency in the latter part of 1929.[7] Intensive development work on similar systems is under way in the United States. The National Broadcasting Company has in operation in its network, two groups of two stations each, which are being operated synchronously using a standard reference frequency transmitted between stations over telephone circuits. The Bell System has developed a common frequency broadcast system using a standard reference frequency suitable for transmission over telephone circuits. This system has been given a practical test in coöperation with the Columbia Broadcasting System. It will shortly be commercially available.

The other method of attack has been to derive the carrier frequency at each station from an independent oscillator. In England,[8,9] electrically driven tuning forks have been used to supply an audio frequency of high stability from which the carrier frequency has been derived by means of frequency multipliers. With this equipment it has been possible to maintain the derived carrier within a few cycles per second of

[3] Frank B. Falknor, "A History of Synchronization," *Citizens Radio Call Book Magazine and Technical Review*, **12**, 38–40; March, 1931.

[4] W. Hahn, *Funk*, **35**, 247–248, 1928.

[5] W. Hahn, *Die Sendung*, **5**, 430–432, 1928.

[6] F. Gerth, "A German Common Frequency Broadcasting System," *Proc. I. R. E.*, **18**, 510–512; March, 1930.

[7] Erik Esting, *Elektroteknik*, pp. 109–112, June 7, 1930.

[8] P. P. Eckersley, "The Operation of Several Broadcasting Stations on the Same Wave-length," *Jour. I. E. E.*, 1929.

[9] P. P. Eckersley, "The Simultaneous Operation of Different Broadcast Stations on the Same Channel," *Proc. I. R. E.*, **19**, 175–194; February, 1931.

isochronism [10] and this has been sufficient to permit a satisfactory service to be rendered to the territories immediately adjacent to each station. As will be shown in detail later there is a substantial difference between the service range of a station operating in almost perfect isochronism with the other stations in the common frequency broadcast system and that of a station which is more than a small fraction of a cycle per second out of isochronism. In this country "matched crystals" and other means of independent frequency control have been tried but the frequency stability of the best equipment available in the past has fallen far short of that required for the satisfactory operation of the stations on a common frequency.

In the spring of 1930 the Central Broadcasting Company of Iowa found itself in the possession of a concrete example of the need for the simultaneous operation of two stations on a common frequency in that its stations WHO and WOC were compelled to divide time equally on 1000 kc. so that the Davenport and Des Moines areas each received service from its local station but half the time. These stations are 153 miles apart and either could be depended upon to render a high-grade service only within a radius of about fifty miles of the station. It was felt that with the simultaneous operation of both stations, each of these areas would receive full time service from its local station.

The Central Broadcasting Company presented its problem and asked for equipment capable of maintaining the carriers of these two stations within the limits of isochronism required for their simultaneous operation. Bell Telephone Laboratories therefore undertook the necessary development work.

The degree of isochronism required for the various conditions existing under the different types of common frequency broadcast systems is in fact a fundamental question that must be answered before any logical delineation of the problem can be attempted. Unfortunately there exists no similar condition in ordinary human experience from which a valid analogy can be drawn, so that the *a priori* assumptions which have been used in the preliminary theoretical discussions of the various phases of this problem have of necessity been based primarily upon personal opinion and the resultant conclusions have quite naturally varied between extremely wide limits.

The problem had been studied intensively during the preliminary field tests of common frequency broadcasting which were made in the fall of 1929 in coöperation with the Columbia Broadcasting System us-

[10] The term "isochronous" has been used instead of "synchronous" in order to exclude the concept of identity of phase which is usually included in the meaning of the latter together with the meaning of identity of frequency which is common to both words.

ing stations WABC and WCAU. It proved to be very difficult to get accurate and consistent data from such field observations without a very extensive series of tests because the fortuitous variations in the transmission medium continually altered the test conditions. These were especially troublesome since the frequency difference between the carriers is but one of the two independent variables of primary importance which affect the quality of the program received at any given point, the other being the ratio of field strength received from the two stations at the point in question.

It was therefore necessary to set up in the laboratory apparatus which would simulate as closely as possible the conditions existing in the field but with all the variables under definite control Two identical miniature transmitters were modulated by the same program. The modulated carriers were then attenuated through independent transmission paths and received by a high-quality detector. The layout of the apparatus is shown schematically in the block diagram of Fig. 1.

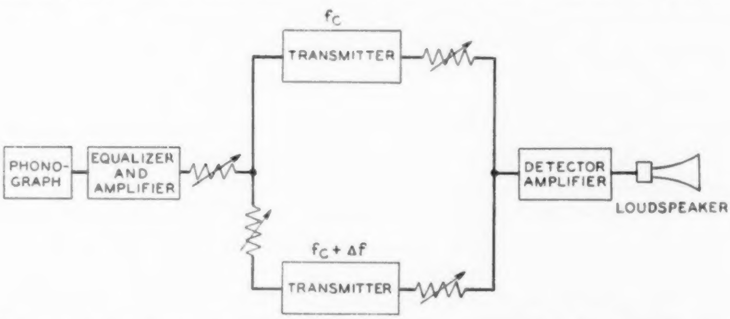


Fig. 1—Block diagram of apparatus set-up for determinations of quality impairment with different degrees of isochronism and field strength ratios.

It will be seen that with this equipment the signal strength received at the detector from either station may be varied independently so that any desired signal strength ratio may be obtained. The frequency difference, Δf, was fixed directly by the adjustment of the carrier frequencies of the two transmitters to the required degree of isochronism. These transmitters, operating at a frequency of approximately 50 kc. were quite stable and capable of accurate adjustment.

The over-all audio-frequency transmission characteristic of the whole system was even better than is available in the better commercial radio receivers. The observers were engineers well acquainted with the effects to be expected and whose judgment was extremely critical. Tests were made with material consisting of both musical and talking

programs and, while the effects are more noticeable with musical programs due to the presence of sustained tones, the difference was not marked. The observers compared the quality of the program received from the two stations with varying field strength ratios and degrees of carrier isochronism with that received from one of the stations transmitting alone. The change from the test condition to the reference condition could be made at will and the gains of the various circuit elements were adjusted so that the apparent program level was the same under the two conditions. Each test covered a considerable period of time and the curves shown in Fig. 2 mark the field strength ratios at

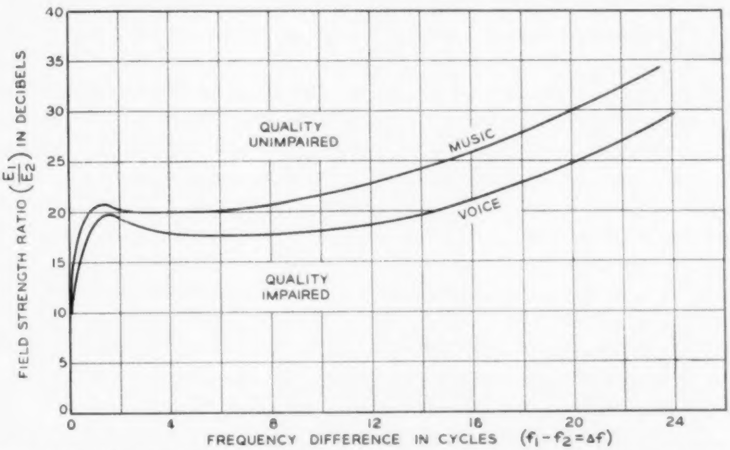


Fig. 2—Quality impairment vs. the frequency difference $f_1 - f_2$ and field strength ratio E_1/E_2. The curves mark points where the quality impairment was just perceptible.

which the observers could not distinguish between the test and reference conditions. The data shown are therefore believed to be distinctly conservative and to represent a criterion much more severe than any which will be encountered in commercial operation. These results are also in agreement with the experimental data that were obtained from our field tests and also check closely the data obtained by the engineers of the British Broadcasting Company in similar field tests in England.

It will be noted that when the frequency difference is very small, closely approaching isochronism, unimpaired reception is assured provided the field strength ratio is at least 10 db but that as soon as the frequency difference is at all appreciable the required field strength ratio for ordinary programs rises sharply to about 20 db and is approximately constant within the range from 1 to 10 cycles per second.

Our field strength distribution surveys and studies have shown that, for 5-kw stations separated by two or three hundred miles, a field strength ratio of 20 db is obtained only at points well within the normal service area of the station. On the other hand the limits of the 10-db ratio lie for the most part outside the normal service range of the station. Thus if such a station is to be operated on a common frequency chain, the carriers must be maintained approximately in isochronism if a large portion of the listeners within the normal service range of the station are not to receive a seriously impaired program. If approximate isochronism is maintained, the service area of each of these stations should not differ materially from that which selective fading and interference would establish for that station transmitting alone.

In order to maintain unimpaired reception in the region where the field strength ratio is between 10 and 20 db, it is necessary that the stations be operated so that their carriers are not permitted to differ in frequency by more than one cycle in 10 seconds and this demands a frequency stability of an entirely new order of magnitude for commercially available independent oscillators. However, at the time that this development was undertaken for the Central Broadcasting Company, previous tests had shown that a newly developed crystal controlled oscillator unit designated as the No. D-90684 oscillator-amplifier possessed an exceptional frequency stability for commercial equipment and that minor modifications would give it the stability required for the simultaneous operation of a small group of stations on a common frequency.

It was therefore planned to replace the existing crystal control equipment by one of these new units located at each station and supplemented by a monitoring receiver located midway between the transmitters.

The No. D-90684 oscillator-amplifier is a relay rack mounted assembly consisting of a shielded unit containing a constant temperature oven and a crystal osciliator, an amplifier having a maximum power output of thirty watts, and the necessary power control equipment. The amplifier tubes, instruments, and controls are mounted on the front of the panels as is shown in Fig. 3 and all other apparatus is mounted in the back and enclosed by a metal locker. The assembly of the various components inside the locker is shown in Fig. 4. The power equipment is placed in the lower part, the constant temperature oven and crystal oscillator unit is mounted on slides in the middle compartment, while the upper shielded section isolates the buffer and output stages from the rest of the transmitter. The door of the locker is fitted with safety switches which automatically disconnect all high voltages from the

equipment before the door can be opened. It was a simple matter to install one of these compact self-contained units adjacent to each transmitter to replace the existing crystal control equipment as the source of the carrier frequency. A corner of the operating room at station WOC is shown in Fig. 5, with a part of the radio transmitter at the extreme right and the oscillator-amplifier mounted adjacent to it. The author is holding the crystal oscillator and constant temperature oven, and over his head to the left is the loud speaker through which the program from the monitoring point is received.

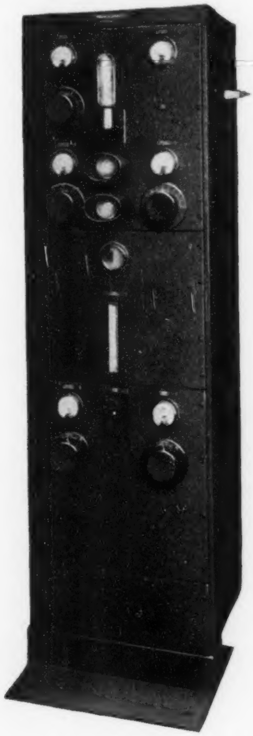


Fig. 3—Front view of crystal controlled oscillator-amplifier unit.

The extraordinary frequency stability of these units has not been obtained through any radical change in design but has come rather as a result of the refinement of all the component elements to form a coördinated unit. A clamped crystal has been used in an improved type of holder, designed to maintain a constant pressure on the crystal

and at the same time to prevent any lateral movement which would cause a change in the crystal frequency. The crystal and its holder are mounted in an oven fitted with an improved thermostat capable of maintaining the temperature of the crystal constant within extremely narrow limits. This constant temperature oven is built as an integral part of the oscillator, which has been designed to work the crystal under the conditions of optimum stability.

The oscillator and crystal are carefully shielded and isolated from the output stage by several buffer stages in order to prevent any change in the load conditions from being reflected back to the oscillator and thereby changing its frequency. Careful tests in the laboratory have shown that the output power could be varied from zero to full load without affecting the frequency within the limits of observation, which

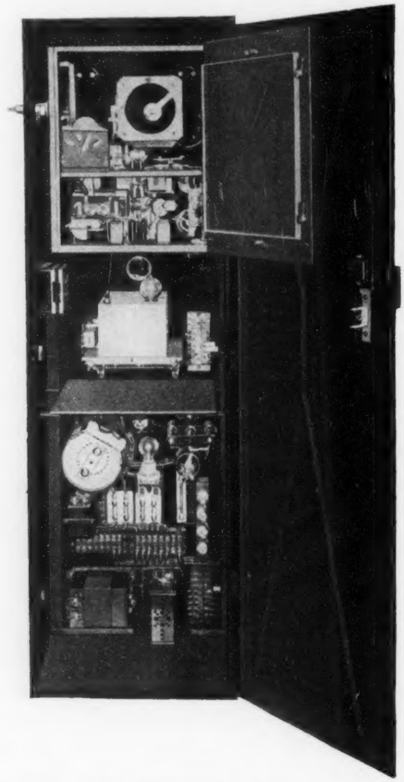


Fig. 4—Rear view showing interior of oscillator-amplifier unit.

were about one part in a hundred million. It is relatively insensitive to changes in filament current, though this is maintained constant within narrow limits by a ballast lamp. Since a change of one per cent

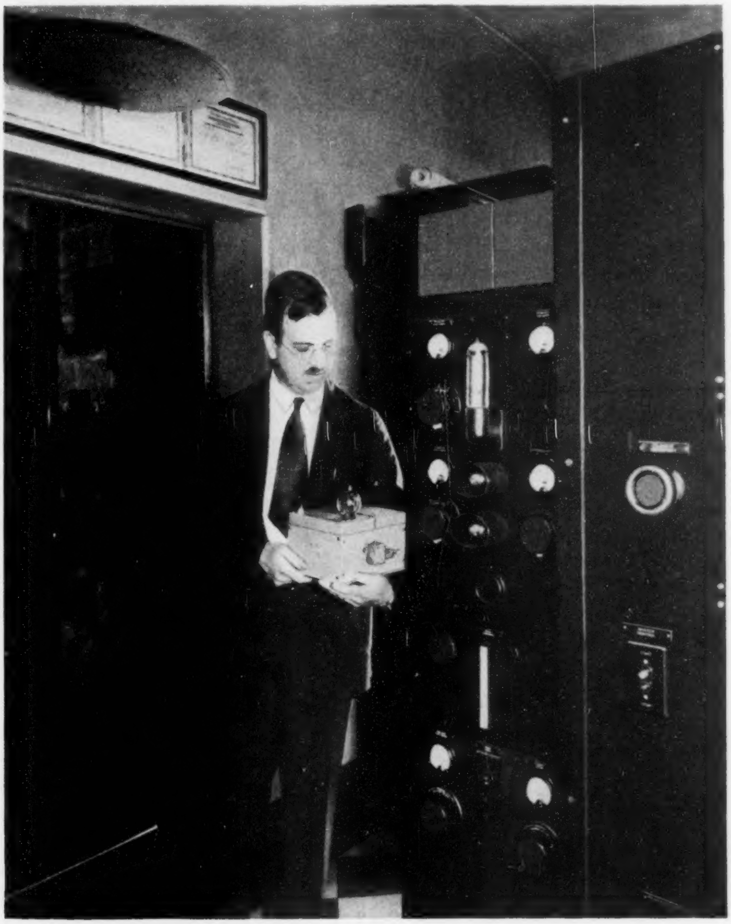


Fig. 5—A corner of the operating room at station WOC.

in the plate voltage causes an immediate change in the frequency of about one part in fifteen million and an ultimate change of about one part in two million, the crystal oscillator is now being operated from batteries.

Since even with these oscillators absolute isochronimc sannot be maintained indefinitely without readjustment, WHO was chosen as the reference frequency station and WOC was provided with means by which its carrier frequency could be brought into exact isochronism with that of WHO. In order that the operator of WOC could easily determine the degree of isochronism, a monitoring receiver was set up at a point midway between the stations and the program received there was transmitted back to station WOC by wire line. A departure of the two stations from isochronism is shown by a slow variation in the level of the program received and the operator can then make the adjustment necessary to restore the stations to isochronism. The nicety of this adjustment can best be appreciated by the fact that a complete revolution of the control dial varies the carrier frequency at WOC by but one part in a million.

It was found difficult at times to determine the beat frequency resulting from a lack of isochronism on account of the masking effect of the rhythm or beat of a musical program. The receiver was therefore equipped with a small audio-frequency oscillator which was arranged to modulate completely the incoming carriers received from the two stations. These combined modulated carriers are detected in the usual way and the output from the receiver is then an audio tone, the level of which is directly proportional to the resultant of the combined carriers. This tone overrides the program and is transmitted back to the station where even very slow changes in its level are easily detected by the operator. This also has the advantage that the degree of isochronism can be determined before any program is broadcast and that any necessary readjustment to restore isochronism can be made during silent periods in the program. This tone is required only at the time of adjustment, and relays at the monitoring point have been arranged for remote control from station WOC by which the audio oscillator can be turned on whenever desired. These relays also permit the operation of either of two receivers and permit the setting of the gain of either receiver at the proper level for day- or nighttime reception. The control panel at the station, shown in Fig. 6, is equipped with supervisory signal lamps which indicate the position of these relays.

The equipment at the monitoring point, shown in Figs. 7 and 8, is mounted on a single relay rack and includes the loop antenna which has been made sufficiently unidirectional to permit the obtaining of an exact balance between the signal strengths received from the two stations. The two radio receivers with their associated audio oscillators and the relay control panel complete the equipment at the monitoring point. The rack is arranged for the complete enclosure of all the equip-

ment by means of dustproof can covers which also serve to prevent any accidental disturbance of the settings and adjustments.

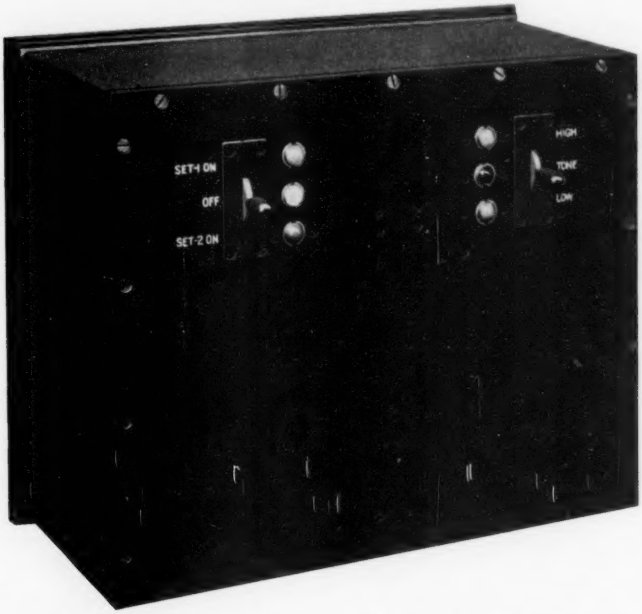


Fig. 6—Control panel for remote control of the monitoring point receivers.

With this equipment in commercial operation, a checking of the frequency every ten minutes in connection with the regular routine inspection of the transmitter has been sufficient to maintain the carriers within an average of two cycles per minute of absolute isochronism. Departures from isochronism of this order of magnitude are not detectable within the normal service area of either station.

While with an installation of this type one is primarily concerned with frequency stability rather than permanence of calibration, the Laboratories have measured the frequency of these stations periodically. It was found at the time of the installation, after the reassembly of the equipment subsequent to its shipment from New York, that the frequency was about two cycles per second different from that measured before shipment. Measurements since that time have shown that the frequency has varied over a period of time between seven cycles above the assigned frequency and seventeen cycles below it. It is known, however, that these variations were primarily due to the

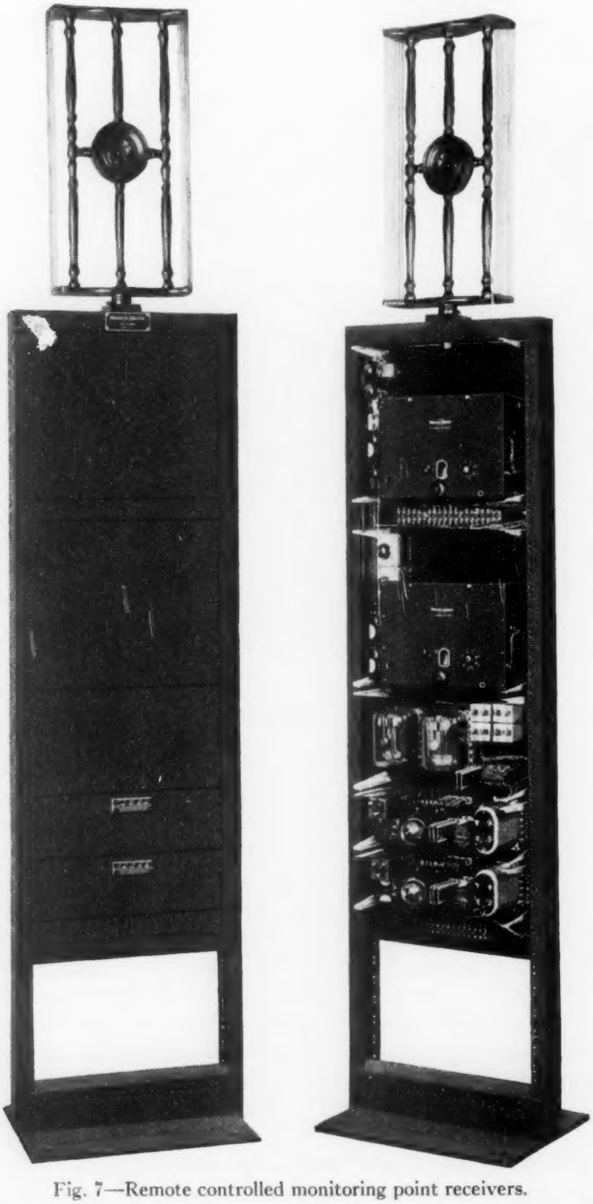


Fig. 7—Remote controlled monitoring point receivers.

Fig. 8—Can covers removed from rear of monitoring point receivers to show the location of the receivers and remote control relays.

variations in plate voltage at WHO, which were permitted in the interest of economy, since the ensuing variations of the two carriers were still far within the requirements of the Federal Radio Commission.

More recent measurements made on a similar unit installed in a broadcast station where precautions have been taken to insure the application of a constant plate voltage to the oscillator have shown a frequency variation of less than five cycles per second from the assigned value over a period of several weeks.

Before approval was sought from the Federal Radio Commission for the full time simultaneous operation of these stations on a common frequency, careful surveys of the areas served were made by the engineers of the Federal Radio Commission, the Department of Commerce, the Central Broadcasting Company, and Bell Telephone Laboratories during their simultaneous operation on an experimental basis during the early morning hours in order to determine the nature of the service being rendered. Nearly three thousand miles were covered by the radio test cars during these tests. Upon completion of these surveys the Federal Radio Commission immediately granted permission for the simultaneous operation of WHO and WOC during regular broadcast hours.

These surveys showed that the service rendered by the simultaneous operation of these two stations was substantially twice as great as the service given on a shared time basis. The normal service area of each station was maintained and the nighttime reception at points over a hundred miles distant from either station was improved by the partial elimination of rapid and selective fading as well as by an increase in the average field strength received.

This improvement in distant reception was confirmed by the letters received in response to requests, made during the tests, for reports as to the quality of reception. In making these requests, the nature of the distortion that might be experienced was carefully described and it was especially emphasized that mere reports of reception would be of no value and that the information desired concerned the quality of the program received during the simultaneous operation of the stations as compared to that from either one alone. Several hundred replies were received from outside the State of Iowa beyond the normal service range of either station. These were almost unanimous in reporting better reception with simultaneous operation. The reports received from distant points during the first year's commercial operation are in full accord with these test data. This improvement apparently occurs wherever marked selective and general fading is experienced in the reception of either station alone.

It has been generally accepted that fading, commonly experienced in the nighttime reception of programs from a distant station, is due to the arrival of the signals along at least two different paths. In the mathematical analysis of this problem it will be convenient to represent each portion of the carrier which arrives at the receiving point via an independent path as a vector of constant amplitude and random phase variation. It will then be possible to represent the fading signal received from a single station as the sum of at least two such vectors. It is then logical to assume that the signal received from two distant stations operating on approximately the same frequency is the summation of at least four of these vectors of constant amplitude and random phase relation. This assumption of random phase relation is valid for any of the common frequency broadcast systems now being developed commercially either here or abroad. If the carriers are derived directly from a reference frequency transmitted via wire line circuits to the several stations, the slight phase variations caused by temperature and humidity changes are sufficient to cause the phases of the derived carriers of the different stations to vary in a fortuitous manner. Furthermore, even if the carriers of two stations were held exactly in phase at their respective antennas, or at some point midway between the transmitters, the variations in the path-lengths of the waves arriving at any given distant point would be sufficient to cause a random phase variation. It is helpful in the mathematical analysis to assume also that these vectors are of equal amplitude. While this is not strictly true in all cases, our field observations have shown that it is the limit which tends to be approached as the distance from the stations is increased.

With these assumptions it can be shown mathematically [10] that the probability P_2, that the ratio of the sum of two vectors to their absolute sum will be less at any instant than a given value λ, is given exactly by the expression

$$P_2 = \frac{2}{\Pi} \sin^{-1} \lambda.$$

For larger values of "n," the exact expression is difficult to evaluate but a close approximation to the probability P_n for "n" vectors is afforded by the expression given below:

$$\begin{aligned} P_n = {} & \frac{12n^2 - 6n + 1}{12n}\lambda^2 - \frac{12n^2 - 18n + 13}{24}\lambda^4 \\ & + \frac{12n^3 - 36n^2 + 55n}{72}\lambda^6 - \frac{12n^4 - 60n^3 + 155n^2}{288}\lambda^8 \\ & + \frac{12n^5 - 90n^4 + 350n^3}{1440}\lambda^{10} - \frac{12n^6 - 126n^5 + 646n^4}{8640}\lambda^{12} \end{aligned}$$

[10] See Lord Rayleigh's, "Scientific Papers," Vol. 6 section on "Flights in 1, 2, and 3 Dimensions," and also section on "Random Unit Vibrations."

This probability of the sum of "n" vectors being less than any given percentage of the absolute sum of "n" vectors has been computed by means of these expressions for the cases corresponding to the distant reception of 1, 2, 3, and 5 stations. The results of these computations have been plotted in Fig. 9.

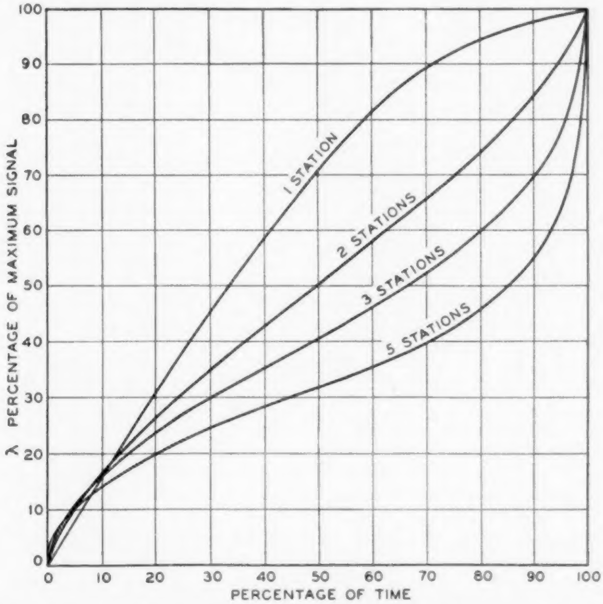


Fig. 9—Curves showing the probability that the instantaneous sum of the signals from a number of distant stations will be less than any given percentage, λ, of their absolute sum.

There are two aspects of these curves which are of especial interest in connection with this problem. First it will be noticed that as the number of stations is increased the percentage of time that the signal fades below a small value such as 5 per cent of its maximum should be decreased. Thus the percentage of time that bad quality will be received due to the elimination of the carrier should be noticeably reduced as the number of stations is increased. Also it will be noted that a rapid reduction in the percentage of time that the signal approaches the maximum should occur as the number of stations is increased. This serves to emphasize the second aspect of the problem, i.e., the level of the signal received should remain near the mean for a much larger percentage of the time as the number of stations is increased. Thus a

distant listener can set his receiver so that a normal level should be obtained for a much larger proportion of the time as the number of stations is increased. As an example, let us consider the proportion of the time that the level of the received program should lie between the limits of 25 and 50 per cent of the maximum signal; for one station it should be but 17.5 per cent while for two stations it should be 32.5 per cent, for three 45 per cent, and for five 55 per cent of the total time. A further development of the probability integral given above has shown that not only should the proportion of the time that a normal program is received increase, but that the instantaneous rate of fading should also de-

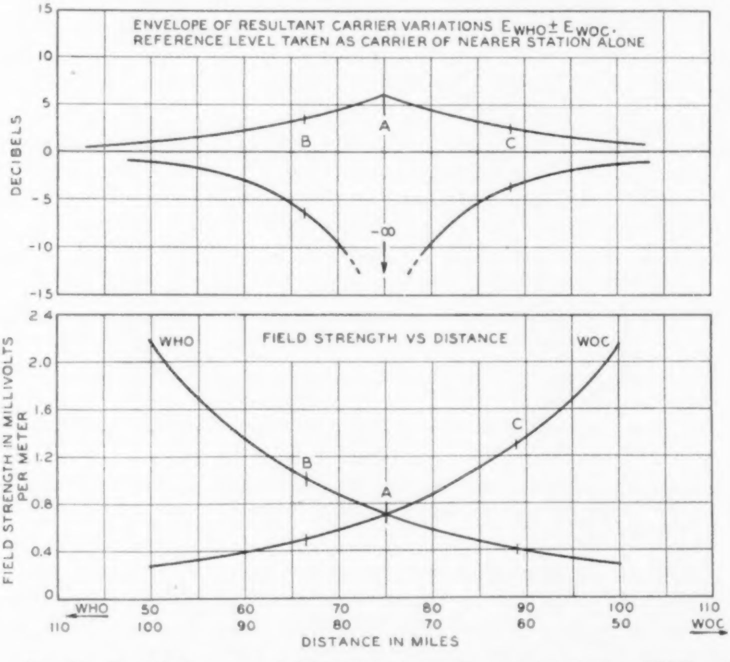


Fig. 10—Smoothed curves of field strength vs. distance for stations WOC and WHO, $fc = 1000$ kc and the envelope of resultant carrier variations for $E_{WOC} \pm E_{WHO}$.

crease as the number of transmitting stations is increased. This is important because the sensitory reaction to fading, within ordinary limits, apparently depends more upon the rate of change of program level than it does upon the absolute total volume change. Since the same arguments apply equally well to each of the individual frequencies comprising the side bands, it can be seen that the general

tendency of increasing the number of isochronously operated stations is to improve markedly the satisfactoriness of the program received at a point distant from all the stations of the chain.

On the other hand in a small area midway between the stations, which received but a mediocre service originally since it lay outside the normal service area of either station, the reception with simultaneous operation was somewhat further impaired. The conditions that exist in this no-man's land between any two stations operating on a common frequency seem worthy of a detailed consideration, especially since wide publicity has been given to the misconception that the maintenance of the carrier in perfect synchronism at the transmitters would entirely eliminate this area of impaired reception. It will be shown below that fundamentally the degree of isochronism merely determines the rate at which alternate strips of bad and good quality reception are swept across this territory. The attainment of exact isochronism would only mean that these strips would tend to be fixed in space and that a certain proportion of the listeners would then receive bad quality all the time instead of getting their share of the good with the bad.

A smoothed curve of the daytime field strengths from WOC and WHO existing in the middle area on a line between the stations is shown in the lower part of Fig. 10.[11] The range of variation that the

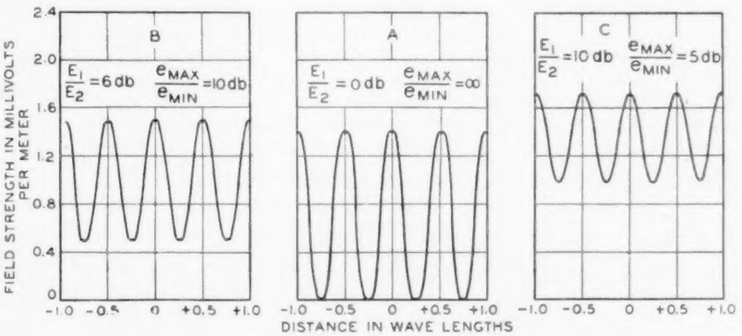


Fig. 11—Enlarged sections of the points *A*, *B*, and *C* of Fig. 10 showing in detail the resultant carrier variations with distance.

resultant carrier level will undergo as the carriers pass in and out of phase is shown in the upper part for the corresponding points along this line. In Fig. 11 enlarged sections showing in detail the variations

[11] These curves are based on field strength measurements made by Radio Inspectors J. M. Sherman and H. T. Gallaher of the Department of Commerce, and furnished through the courtesy of Mr. H. D. Hayes, U. S. Supervisor of Radio, Chicago, Ill.

of the resultant carrier level with distance are given for the point of equal signal strength A and for the points B and C where the field strength ratio is 6 and 10 db, respectively. Any departure from isochronism will have the effect of making these points of maxima and minima move along this line in space at the rate of one-half wave-length per cycle difference in frequency.

Now since the side band frequencies must perforce differ in wavelength from the carrier, they will arrive at any given point out of phase with the carrier, the amount depending on the distance from the transmitter and the side band frequency. Thus the side bands will not for the most part be in phase opposition at the same points in space as are the carriers, and distortion will result from the elimination of the carrier while strong side band components are present. The magnitude of this distortion is primarily a function of the existing field strength ratio between carriers and, while the distortion occurs for only a small proportion of the fading cycle, it is extremely objectionable where the field strengths approach equality. Here the carrier is almost entirely eliminated momentarily and the resultant program consists mainly of second harmonics and other distortion products.

It is entirely outside the scope of this paper to attempt to present a complete analysis of this problem but an effort has been made to indicate the quantitative results that may be expected by selecting a few typical examples. The signal being detected has been assumed in all cases to consist of the ordinary carrier and double side band transmission. The theoretical work which follows has been based upon the use of a square-law detector as being representative of the majority of the existing receivers. In order to avoid undue complexity the curves have been computed for a single frequency audio signal.

In a square-law detector distortion appears primarily in the form of second harmonics and the ratio of these to the fundamental has been taken as a measure of the distortion present under the varying conditions of reception that may exist in the middle area between the stations. There are so many variables concerned in this problem that it is necessary to hold first one and then another fixed while different aspects of the situation are studied.

The first set of curves, Figs. 12, 13, and 14, shows the conditions which exist at the point directly between and equidistant from the two stations when the audio signal supplied to the two transmitters is exactly synchronized, i.e., the audio phase angle $\beta = 0$. With this variable fixed the curves in each successive figure of the series have been plotted for successively decreasing signal strength ratios in order to show the effect of the varying radio phase angle with different degrees

of modulation. It will be seen from these curves that making the modulation of the two carriers equal effects a tremendous reduction in the amount of distortion present in the hollows of the fading cycle that occur when the carriers approach phase opposition, i.e., $\gamma = 180$ degrees. Also it will be seen from a comparison of the family of curves for 100 per cent and 50 per cent modulation ($M_2 = 1$ and $M_2 = 0.5$),

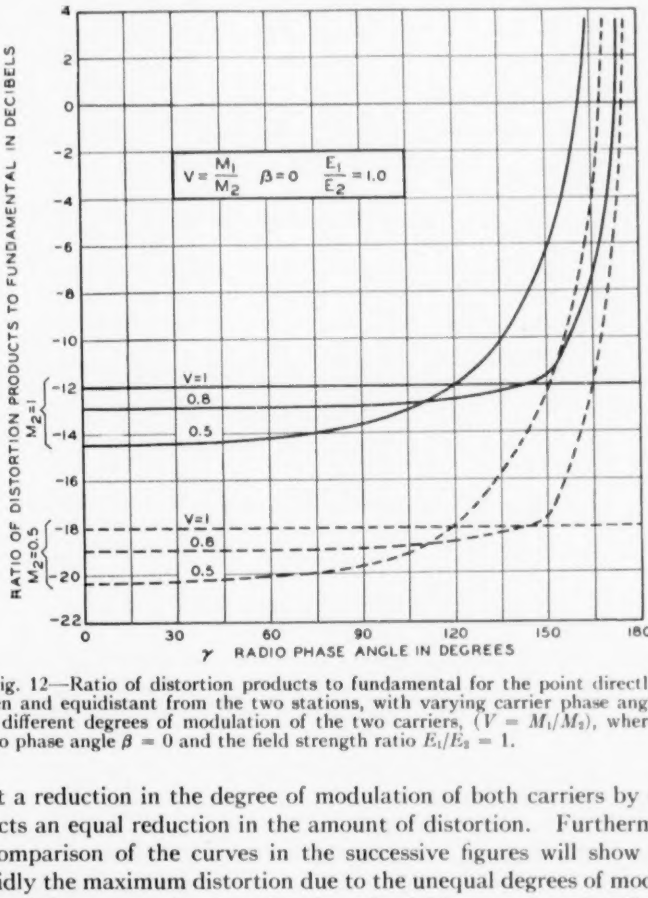


Fig. 12—Ratio of distortion products to fundamental for the point directly between and equidistant from the two stations, with varying carrier phase angle, γ, and different degrees of modulation of the two carriers, ($V = M_1/M_2$), where the audio phase angle $\beta = 0$ and the field strength ratio $E_1/E_2 = 1$.

that a reduction in the degree of modulation of both carriers by 6-db effects an equal reduction in the amount of distortion. Furthermore, a comparison of the curves in the successive figures will show how rapidly the maximum distortion due to the unequal degrees of modulation of the two carriers is reduced as the field strength ratios diverge from unity.

In the second series of curves, Figs. 15, 16, and 17, the effect of varying the carrier phase angle is shown for different representative

values of audio phase angle while the degree of modulation of the carriers is fixed and equal and the field strength ratio is given a different

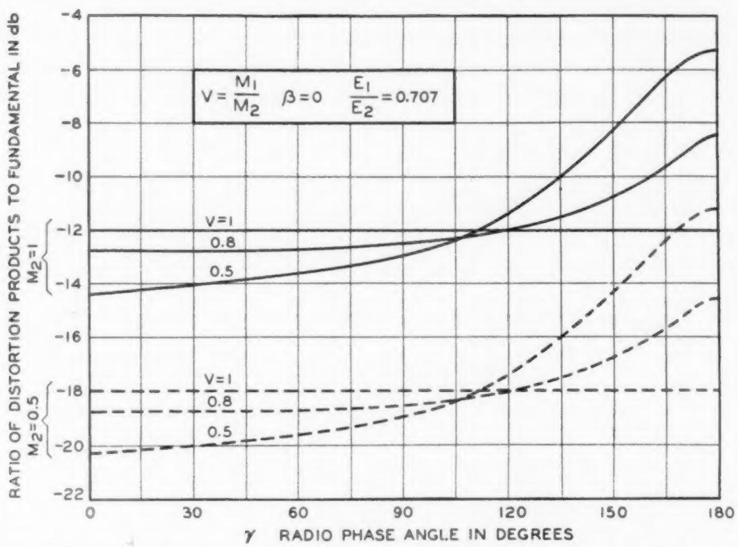


Fig. 13—Same as for Fig. 12 except that the field strength ratio $E_1/E_2 = 0.707$.

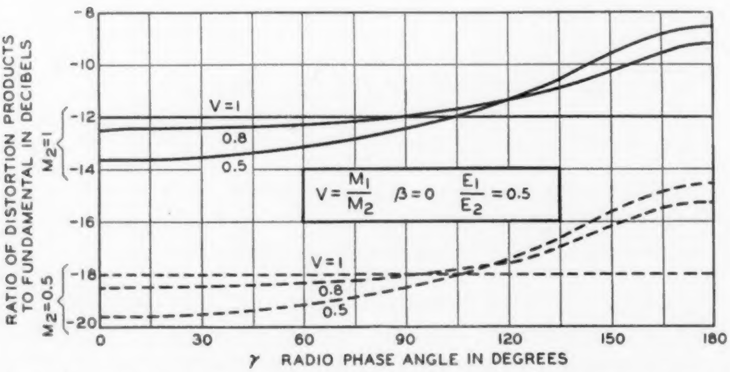


Fig. 14—Same as for Fig. 12 except that the field strength ratio $E_1/E_2 = 0.5$.

value in each successive figure. Here the marked increase in the amount of distortion present as the audio components depart from synchronism and the carriers approach phase opposition is most strik-

ing. Also the rapid decrease in the amount of distortion present as the field strength ratio diverges from unity is noteworthy where $\beta \neq 0$.

The distortion for values of $\beta = 0$ and equal degrees of modulation ($V = 1$) remains constant in both these series of curves because this is the limiting case and is the distortion that would result from the re-

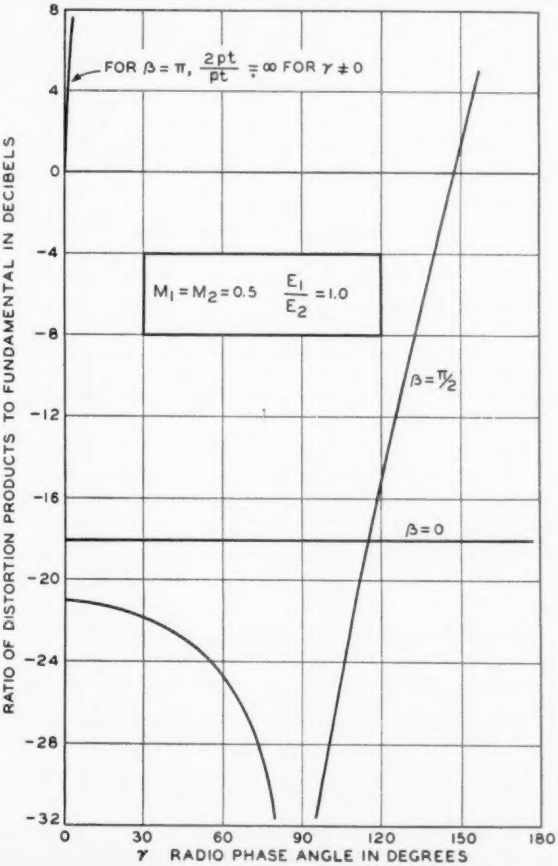


Fig. 15—Ratio of distortion products to fundamental with 50 per cent modulation of each carrier for varying audio and radio phase angles, β and γ respectively, where the field strength ratio $E_1/E_2 = 1$.

ception of a similar program from but a single station. This fact affords a basis for comparison in considering the additional distortion that results under certain conditions from the simultaneous operation

of two stations. While the distortion products loom large in proportion to the fundamental at times, these are also the times when the fundamental is fading out and the actual magnitude of the distortion products is not large.

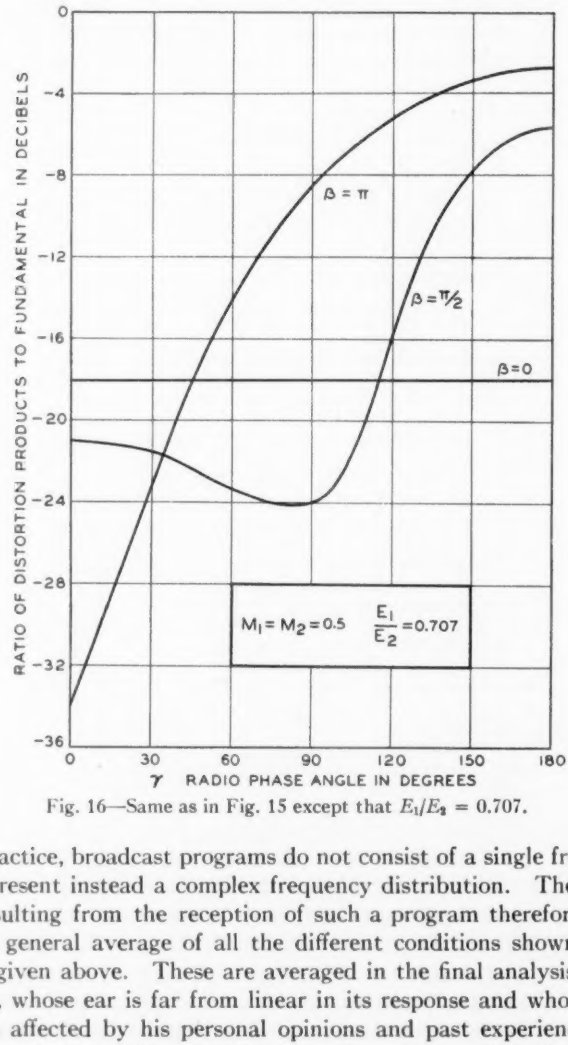


Fig. 16—Same as in Fig. 15 except that $E_1/E_2 = 0.707$.

In practice, broadcast programs do not consist of a single frequency but represent instead a complex frequency distribution. The distortion resulting from the reception of such a program therefore represents a general average of all the different conditions shown in the curves given above. These are averaged in the final analysis by the listener, whose ear is far from linear in its response and whose judgment is affected by his personal opinions and past experience. We have therefore considered it futile and perhaps misleading to at-

tempt to present any graphical summation of the effective distortion present in the reception of an ordinary broadcast program under such conditions. On the other hand these theoretical studies were undertaken in order to explain certain phenomena observed during the preliminary field tests as the degree of modulation and the field strength ratio were varied. The actual results have been quite closely corroborated by the conclusions reached from a study of these single frequency curves which have been of great value in obtaining a physical picture of the conditions that exist in this middle area.

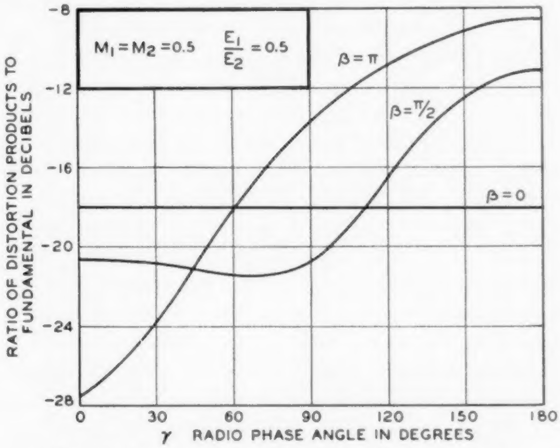


Fig. 17—Same as in Fig. 15 except that $E_1/E_2 = 0.5$.

These curves have been limited, for the sake of simplicity, to the consideration of the conditions at the points in the middle area equidistant from the two stations. It will be seen from these curves for this limited case that if, with equal degrees of modulation, each individual frequency component of the program could be synchronized at the two transmitters no additional distortion would be caused at these points by the isochronous operation of the two stations. But for points not equidistant the audio phase angle will not be zero even though it is maintained so at the transmitter and the magnitude of this divergence from synchronism will be different for each audio frequency and for each separate point in space. The magnitude of this divergence increases rapidly as the distance to the respective stations becomes more unequal. Furthermore, the problem of maintaining in synchronism every component of the broad frequency spectrum required for program transmission appears to offer tremendous technical difficulties.

It seems especially questionable whether such synchronism is necessary since tests in this area have shown that the use of a simple directive antenna capable of moderate discrimination against the weaker of the stations at the point in question is sufficient to render the reception at least comparable to that from either station alone. A loop antenna grounded at one side instead of the center was found to be very effective.

Population studies made in connection with the field surveys show clearly how marked is the improvement in the service rendered by these particular stations under simultaneous operation as compared with operation on a shared time basis. On a shared time basis a population of approximately 1,000,000 received adequate service from these stations but half the time, the value of which was greatly impaired by its intermittent character. With simultaneous operation the service area of each station receives full time service. No accurate estimate can be made of the number of people in the middle area, and outside the normal service range of either station alone, whose reception has been further impaired by simultaneous operation. The importance of this effect can, however, be estimated from the fact that but 60 complaints of impaired reception were received by these stations in the first 35 days of simultaneous operation during the regular hours and that the total for the first year is less than one hundred.

The marked increase in the service rendered by these stations through simultaneous operation is an indication of the possibilities of the improved service that can be made available to urban areas by the use of isochronized transmitters for the broadcasting of a common program. Although it is probable that the high powered station on a cleared channel will remain the best means of affording a high-grade service to a metropolitan area while also rendering an acceptable service to large rural areas, common frequency broadcasting now appears to offer a definitely useful means by which to provide an improved coverage to a number of noncontiguous communities.

In conclusion, the author wishes to acknowledge his especial indebtedness to the following members of Bell Telephone Laboratories: to Mr. G. R. Stiblitz for the development of the probability curves, and to Mr. C. B. Aiken and Mr. R. J. Jones for their preparation of the distortion curves as a part of their general mathematical study of the problem.

Application of Printing Telegraph to Long-Wave Radio Circuits *

By AUSTIN BAILEY and T. A. McCANN

This paper describes certain arrangements which have been used for start-stop printing telegraph operation over a transatlantic long-wave radio channel and also describes results obtained from certain tests of long-wave teletypewriter transmission from Rocky Point, L. I., to Rochester, N. Y. A prediction of year-round results is obtainable by correlation of these test data with year-round noise measurement data taken at Houlton, Maine, in connection with transatlantic telephone service.

PRINTING telegraph equipment,[1] because of its speed, accuracy and convenience in transmitting intelligence, has become recognized as a very useful method of telegraphy on wire circuits. It seems important, therefore, to determine something of the possible utility of present types of teletypewriters on radio circuits.

It is common practice to transmit the signals for operating teletypewriter equipment over wire circuits in any one of several electrical forms. As in earlier telegraph practice the signals are frequently transmitted as d-c impulses. More recently alternating currents of voice-frequency and of higher frequency have been employed.[2] In employing radio frequencies for operating teletypewriter equipment where the operating impulses are no longer guided by a wire circuit, new problems and new conditions arise, which are essentially those of radio telegraph transmission. For this reason, it is desirable to review briefly the conditions under which radio telegraph systems are operated.

In manual-sending aural-receiving practice for radio telegraphy it has been customary to utilize, at the receiving end, only a marking tone or sound which is received during intervals corresponding to the time that the sending key is depressed. In transmitting signals from an arc transmitter, a signal of a different frequency is sent out during the spacing periods in order to simplify the keying process, but this spacing signal is not utilized at the receiving end. For aural reception the necessary and sufficient requirement is that the marking tone be distinguishable through the noise. Using ear receiving it is possible to distinguish the signal under a wide variation of conditions because of the ability of the ear to accomodate itself to variations in signal level and in signal-to-noise ratio.

* Presented at Sixth Annual Convention of I. R. E., Chicago, Illinois, June 4–6, 1931.

From the transmission standpoint, tape or automatic sending, i.e., sending from a tape perforated in accordance with a telegraphic code, is merely a matter of increasing speed and accuracy of the characters transmitted.[3] Automatic tape recording of radio signals, such as by the syphon recorder or similar device, removes the advantage obtained from the tone character of the signal, so useful in ear reception, and substitutes for this tonal character the less acute ability of the eye to distinguish between signals and noise on the tape record.[4] In sacrificing this ability to receive with greater accuracy in the presence of considerable noise there is, however, a gain in the speed of receiving radio signals. The tape record, which is in permanent form, makes it possible for several operators simultaneously to transcribe different parts of the received message at speeds much slower than the transmitting speed.

Printing telegraphy goes one step further in removing the human element from the process of receiving and substituting a mechanism which must be impelled to a definite act by each current element received. The printing mechanism inevitably records what it receives without using any judgment factor in the process other than the mechanical application of such fixed criteria as have been put into it by the designer. Unless the transmitted signal is received with such intensity and character as to be the controlling signal at the receiving end, errors will usually result. The use of printing telegraph equipment on radio circuits,[5] therefore, makes the signal-to-noise ratio necessary for the receiving of satisfactory copy greater than would be required for either aural reception or tape signal recording.

It is of considerable interest to compare the approximate minimum values of signal-to-noise ratio required for satisfactory * transmission of intelligence by single side band long-wave radio using the customary double side band carrier telegraph. This has been done in the table on the following page.

When automatic means for recording the signals are applied at the receiving end of a radio circuit it is desirable that considerable uniformity exist in the output level of the receiving equipment. This is even more important when printing equipment is used. Such a condi-

* Obviously, "satisfactory" cannot have a definite quantitative meaning which is applicable to all modes of communication under all variations in the observed types of received noise. For example, "crashy static" would probably not be as serious in receiving by ear as it would be in receiving by other means. Then too, there is the personal judgment factor in determining just what constitutes "satisfactory" communicaton. The table is set up on a relative basis using quantitative values of signal-to-noise ratio which appear to represent the worst condition under which communication could be effected with only an occasional error. Of course, communication can be continued under much worse conditions, but with an increase in the number of errors.

TABLE I

Type of Facility	Speed of Transmission (Words per Minute)	Approximate Radio Band Width Occupied (Cycles per Second)	Approximate Minimum Signal-to-Noise Ratio for Satisfactory Communication ($20 \log_{10} S/N$)**
Manual-sending, aural-receiving. cw	20	35	10
Manual-sending, aural-receiving. cw	30	50	15
Automatic-sending tape-recording. cw	80	140	20
Single-tone printer system	60	110	30
Two-tone printer system.	60	220	30
Single side band telephony	200	2700	40***

** N is assumed to be measured in a constant band width of about 2200 cycles using the "warbler method" [6] of noise measurement.

*** S is assumed to be about 5 db above 1 milliwatt where speech is at reference volume.

tion is rather to be expected inasmuch as ultimately in the system there must be a relay mechanism operated by the signals. This relay must with a certain degree of accuracy reproduce the length of the signal impulse. It is desirable to have the relay remain unbiased over a considerable range of variation in signal level. If a signal impulse is transmitted only for marking, the spacing signal becomes an interval of no current and the restoring force on the relay must be applied locally by either electrical or mechanical means. Then with signals of the usual rounded wave shape, if the relay operating force varies while the restoring force remains constant, the signals are either "heavy" or "light," that is, the marking intervals are either lengthened or shortened and the system becomes biased.[7] *

The most obvious way of avoiding these difficulties is some arrangement in which the restoring force on the relay is varied in a manner similar to the operating force resulting from the received signal. One method of accomplishing this result which has been found quite effective is the two-tone method of transmission. As far as the radio circuit is concerned the signals consist of a marking and a spacing signal transmitted on slightly different frequencies. Since these two signals traverse the same transmission medium, they are, at least when there is no selective fading, subjected to similar variations in the equivalent of the transmission path. Therefore, if a polar receiving relay is operated by using one of these frequencies to produce the operating force and the other frequency to produce the restoring force, no bias results. The increase in magnitude of variations of the transmission

* Details are given on this effect and methods for its measurement in reference 7.

circuit which can be tolerated by employing this two-tone method of transmission instead of the single-tone method with a fixed bias is shown by Fig. 1.

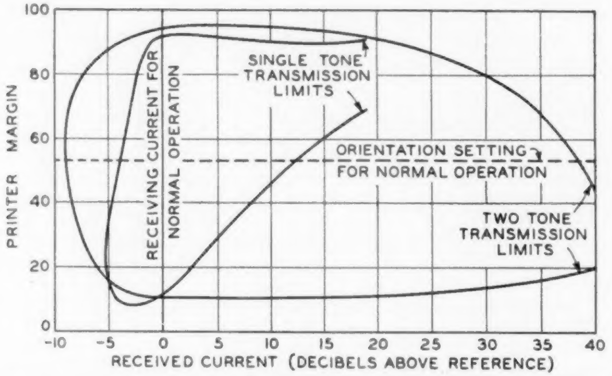


Fig. 1—Relation between received current and printer operating margin for the two-tone and single interrupted tone methods of signaling.

These curves show the relation between operating current and the limits of printer margin [7] within which correct operation is secured for both the single interrupted tone method of signaling and for the two-tone method. The upper and lower limits of the printer margin are shown to meet at the lower levels of received current, indicating that the printer fails completely at these levels. As might be expected, increasing the received current level a few db does not affect the orientation range as seriously as a corresponding decrease. In each condition, however, the margin is less affected when signaling by the two-tone method. Were it not for the presence of noise on the radio circuit, it would be possible to establish the normal operating current at a higher value. The reason why this is not feasible is that the detectors in the voice-frequency telegraph receiving equipment are operated near the upper bend of their characteristics. Under this condition increasing the gain causes a relatively small increase in current from the rectifier that is receiving both noise and signal inputs while the current from the rectifier that is receiving noise only is increased.

Thus because of the desirability of operating through high noise levels on long-wave radio circuits, it is not advantageous to utilize all of the available protection against signal level changes. Rather, a compromise is sought which will afford satisfactory protection against reasonable signal level variations without making the receiving equipment unduly vulnerable to noise. This practical operating point has

been selected at the zero indicated on this figure. The tolerance in received current level variations usually obtained is about ± 3 db in the case of single-tone signaling as compared to about ± 7 db for two-tone signaling. On the transatlantic long-wave radio circuits variations greater than those tolerated by the two-tone transmission method seldom occur with sufficient rapidity to escape manual correction.

With the two-tone system the amount of noise entering the receiving mechanism comes in through double the band width used in the single-tone system,[8] and the intelligence transmitted is completely contained in both the marking and the spacing signals. It is, therefore, logical to expect that there will not be much difference between the two-tone and single-tone systems from the noise interference standpoint. If there were no received signal the noise through the marking and spacing filters probably would balance out to some extent but during operation either the mark or space signal is always present. The noise may effectively annul either signal by being approximately of equal intensity and opposite phase, but the noise through the other filter is received with the full intensity and, therefore, may operate the relay falsely.

Employing printing telegraph equipment on radio circuits is not new.[5,9,10,11] There has, however, been comparatively little commercial use of such systems and there have been very few quantitative data published. Such practical information and quantitative data as have been obtained by the Bell System regarding the application of printing telegraph to radio circuits relate to long-distance overseas point-to-point communication and a short-distance point-to-point overland circuit. Both of these circuits were operated on long waves (about 60 kilocycles).

For the past three years printing telegraph has been employed on the long-wave radio telephone circuit [12] between New York and London to exchange information pertaining to the operation of this telephone service. The printer is admirably suited to this kind of service since the information exchanged frequently consists of foreign names of places and people not familiar to the switchboard operators. By the use of the printer, these can be spelled out with speed and accuracy without the necessity of attempted pronunciation.

The printing telegraph arrangements provided at New York for use of the telephone traffic department on the transatlantic circuits are shown in Fig. 2. The instruments are installed on the table in the foreground. This table is located just behind the switchboard operators. As a large majority of the business transacted is of a question and answer nature, there are special arrangements in the printer to

indicate whether the message printed originated with the New York or with the London operator. Messages transmitted from the local machine are typed in red while those received from the distant terminal are typed in black. This was accomplished by modifying the mechanism of the machine to automatically shift a half red and half black typing ribbon.

Fig. 2—Transatlantic telephone operator's position showing arrangement of printing telegraph equipment.

The voice-frequency telegraph terminal equipment [2] and its associated apparatus are shown in Fig. 3. The equipment comprises the voice-frequency terminal set for repeating between the local d-c printer loop circuit and the a-c line circuit. The printer switching circuits, testing arrangements, and monitoring equipment are, for convenience, included in the same assembly of apparatus. The installation includes all the equipment necessary for one channel of a two-tone carrier tele-

graph system and sufficient equipment for adding another by providing suitable filters and a small amount of additional apparatus.

The connection of the printers to the telephone circuit is shown schematically in Fig. 4. The transmitting telegraph circuits are not

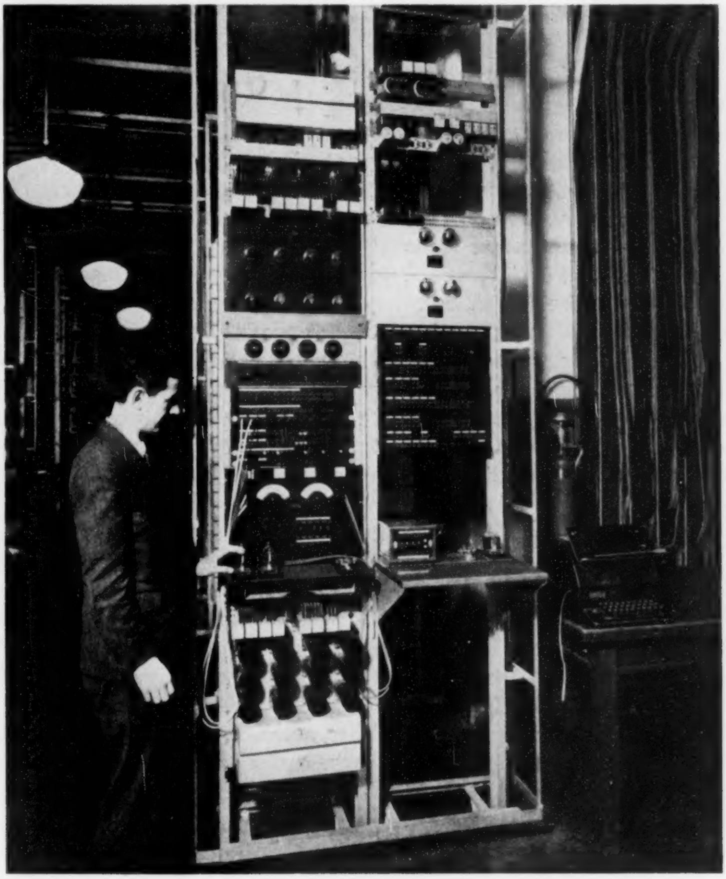


Fig. 3—Equipment for use in applying printing telegraph signals to the transatlantic radio telephone circuit at the technical operator's position.

connected permanently to the radio channel. When it is desired to establish the telegraph circuit, the connection is made through the operator's cord circuits into the transatlantic two-wire telephone circuit in a manner similar to that used to connect telephone subscribers. Audible

monitoring arrangements are provided for the telephone operator, the technical operator, and the printer operator. The distant terminal operator may interrupt the printer circuit with voice if such interruption seems expedient. In addition to the printer used by the printer operator, a printer at the technical operator's position, not shown in Fig. 4, is continuously connected to monitor on the system.

It should be noted that in the arrangement shown in Fig. 4 two different voice frequencies are used for transmission and two others for reception, thus giving the advantages of the two-tone method of transmission. The voice-frequency tones go out over wire circuits to the transmitting station at Rocky Point where, by means of the single side band suppressed-carrier method of radio transmission [13,14] shown in Fig. 4, they are changed to radio frequencies of about 60 kilocycles and amplified. The equivalent radiated power for each frequency is about 50 kilowatts. For signals coming from England much the same process is followed at the British end, the radiated frequencies, however, being different from those transmitted in the opposite direction.

The most serious handicap in the use of printing telegraph on the long-wave channel is noise. During the winter months little trouble is experienced, but interruptions are frequent and are occasionally of several hours' duration during the summer months. At the radio receiving stations the directive antenna systems used for telephony [15] greatly reduce the noise received.

Another important factor in reducing receiving interference is the frequency selectivity. The radio receiver itself restricts the received band sharply to that required for single side band telephony, passing a band about 3000 cycles wide. This is accomplished by the single side band carrier resupplied receiver [14] shown in Fig. 4. The band admitted to each of the tone channel detectors beyond the receiver is narrowed down to about 110 cycles by a voice-frequency filter as indicated in Fig. 4. It is estimated that if the printer were used continuously on the long-wave transatlantic channel for the entire year, the per cent of errors would exceed 0.1 per cent less than 12 per cent of the time and 5.0 per cent less than 2 per cent of the time.

In order to obtain more accurate quantitative information regarding the effect of noise on the transmission of teletypewriter signals over radio circuits, a series of tests was carried out during 1930. For the purpose of these tests a radio circuit was established between the transmitting station at Rocky Point, L. I., and a temporary receiving station at Rochester N. Y., a distance of 286 miles.

This one-way circuit utilized at New York the transatlantic transmitting facilities for printing telegraph described above with the ex-

the
inal
rup-
nter
n in

two
s for
ans-
the
side
Fig.
am-
bout
cess
eing

the
le is
y of
re-
ny[15]

the
ived
ng a
side
tted
wed
Fig.
the
f er-
l 5.0

ard-
over
the
ans-
tion

ans-
ex-

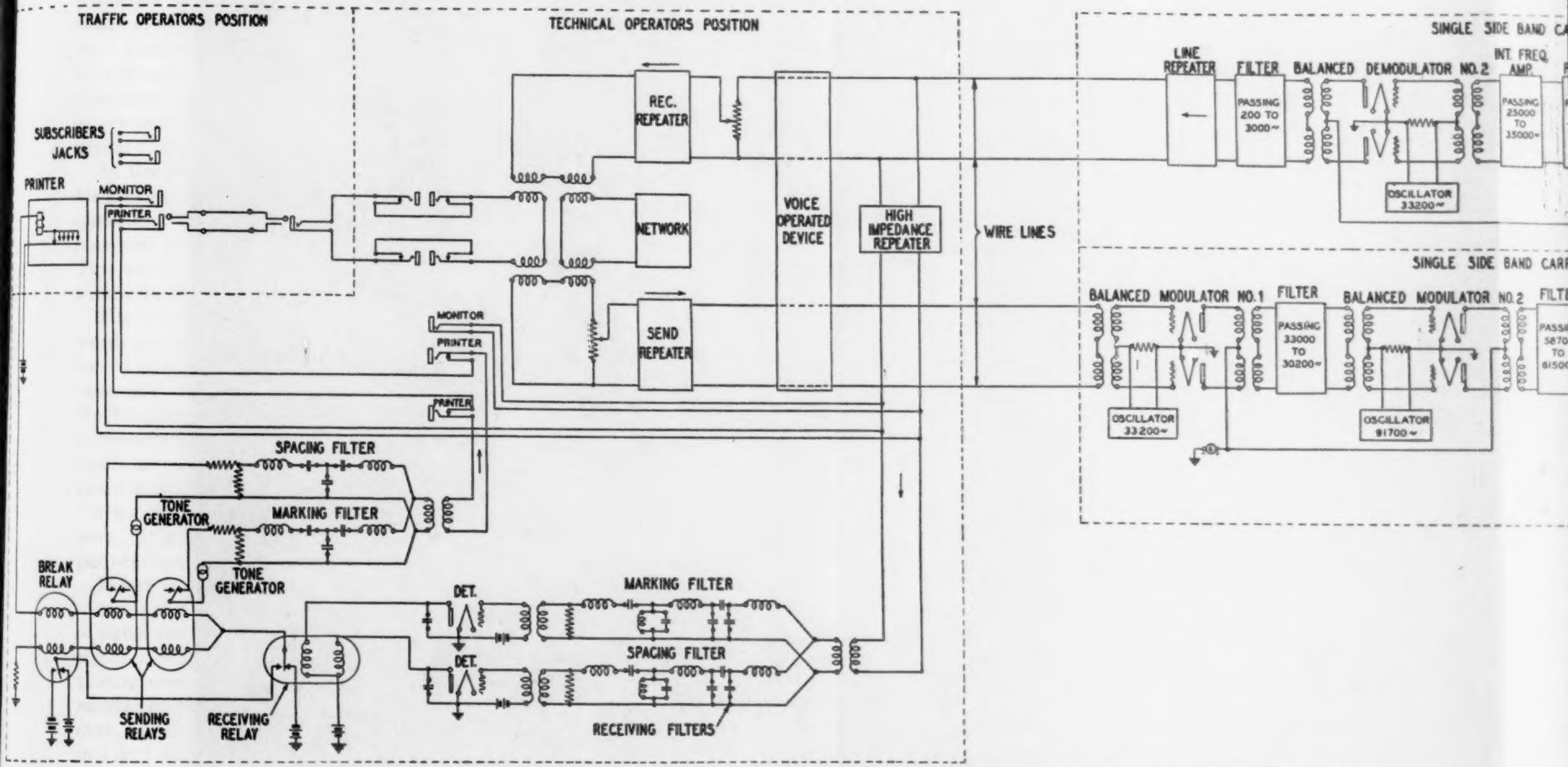


Fig. 4—Schematic of printer connection to transatlantic radio telephone circuit.

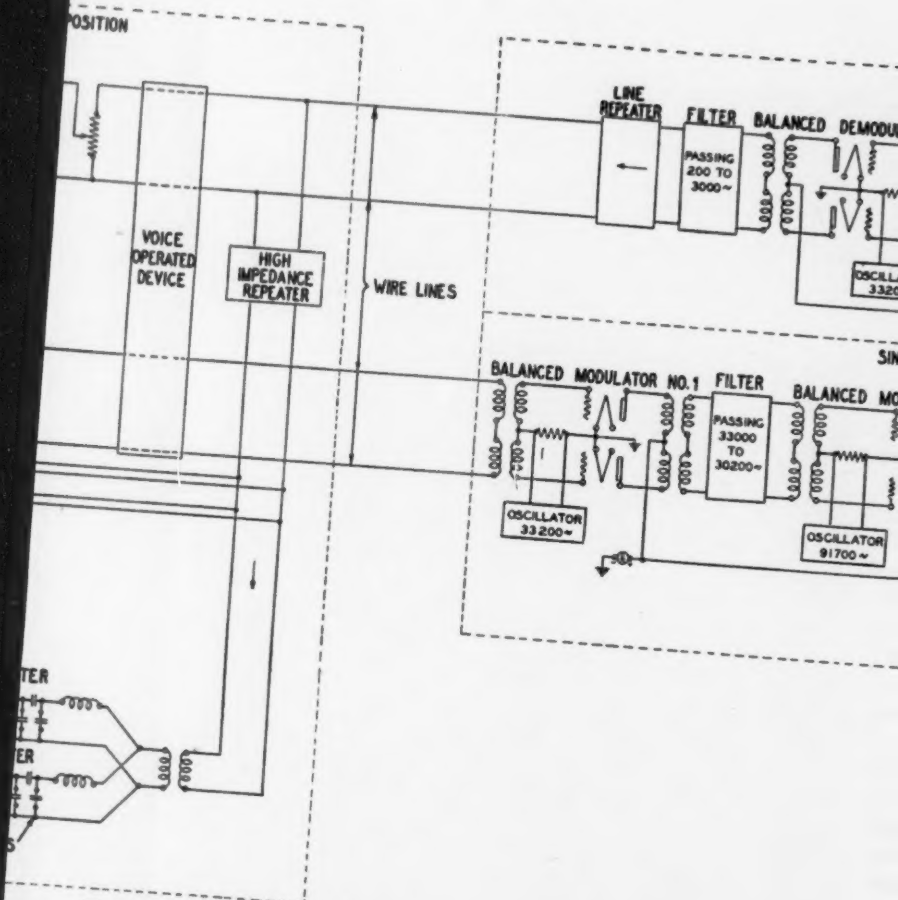
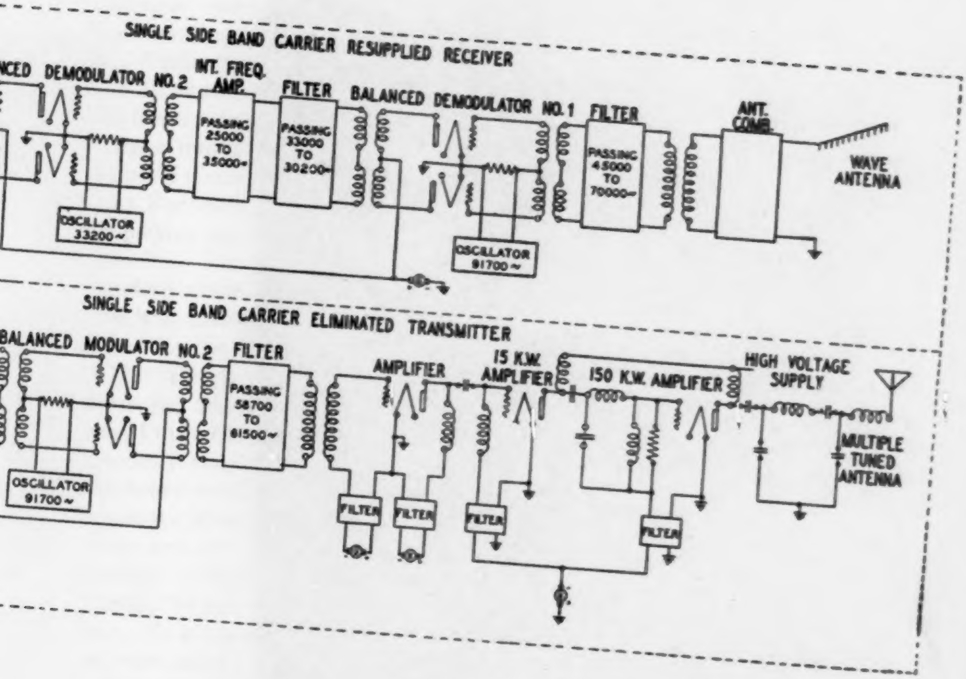


Fig. 4—Schematic of printer connection to transatlantic radio telephone circuit.


SINGLE SIDE BAND CARRIER RESUPPLIED RECEIVER
NCED DEMODULATOR NO. 2
INT. FREQ. AMP.
PASSING 25000 TO 35000~
FILTER
PASSING 33000 TO 30200~
BALANCED DEMODULATOR NO. 1
FILTER
PASSING 45000 TO 70000~
ANT. COMB.
WAVE ANTENNA
OSCILLATOR 33200~
OSCILLATOR 91700~
SINGLE SIDE BAND CARRIER ELIMINATED TRANSMITTER
BALANCED MODULATOR NO. 2
FILTER
PASSING 58700 TO 61500~
AMPLIFIER
15 K.W. AMPLIFIER
150 K.W. AMPLIFIER
HIGH VOLTAGE SUPPLY
MULTIPLE TUNED ANTENNA
OSCILLATOR 91700~
FILTER
FILTER
FILTER
FILTER

cept
ing
meth
A
redu
Roc
tion
db.
42.5
this
in t
peri
A
for
que
stan
This
show
T
wor
ceiv
tane
Key
auto
erro
poss
Dur
sent
per
B
of er
of si
sign
taini
mea
ever
in F
A
time
syst
tic t
data

ception that automatic transmission from a perforated tape for operating page teletypewriters was substituted for the manual keyboard method for operating tape printers ordinarily used.

At Rocky Point the power of the radio transmitter was greatly reduced for these tests. The average power radiated in the direction of Rochester was equivalent to 0.7 kilowatt radiated from a nondirectional antenna. The average deviation from this value was less than 1 db. Under these conditions the average field received in Rochester was 42.5 db above one microvolt per meter. The average deviation from this mean value was less than 2 db. A daily half-hour test was made in the afternoon or evening at a time so chosen as to avoid the sunset period of disturbed radio transmission.

At Rochester, laboratory type receiving equipment was employed for picking up the radio signals and demodulating them to voice frequencies. The voice-frequency signals were then used to operate standard voice-frequency carrier terminal equipment at Rochester. This was modified for two-tone operation in a manner similar to that shown for the transatlantic receiving terminal in Fig. 4.

The teletypewriter signals were sent out from New York at 60 words per minute from an automatic tape transmitter. The copy received over the radio circuit was subsequently compared with simultaneously recorded copy which was not sent over the radio circuit. Keyboard errors which occur occasionally in perforating the tape for automatic transmission appeared on both copies. Disregarding these errors and counting all which did not appear on both copies, it was possible to obtain the per cent of errors caused by radio transmission. During the half-hour daily test period, about 10,000 characters were sent. It is apparent that rates of error which were less than about 0.1 per cent could not be determined accurately.

Before making the half-hour test each day to determine the per cent of errors received at Rochester, measurements were made of the amount of signal and of noise in the output of each voice-frequency filter. The signal-to-noise ratio thus measured was assumed to be the value obtaining over the succeeding half hour of test. The nature of these measurements was such that the data were somewhat scattered. However, by suitable smoothing procedures the approximate curve shown in Fig. 5 was plotted.

At Houlton, Maine, routine radio noise [6] observations are made four times each day on a loop antenna and hourly on the wave-antenna system, as a part of the operating procedure in maintaining transatlantic telephone service.[15] It seemed desirable to find out whether these data which extend over several years could be utilized to extrapolate

the Rochester data into other months. An examination of the noise data observed on the loop antenna at Houlton along with the loop antenna received noise obtained at Rochester, New York, point by point during the period of these tests indicated a fairly constant difference between the noise at these two places. On 37 days during September, October and November 1930, observations of printer operation at Rochester and noise observations at Houlton were made within the same hour. Using the errors observed in the Rochester radio copy on these 37 days and the corresponding 37 values of loop noise at Houlton the cumulative curves shown in Fig. 6 were obtained. From these two curves the same relation as shown in Fig. 5 can again be obtained.*

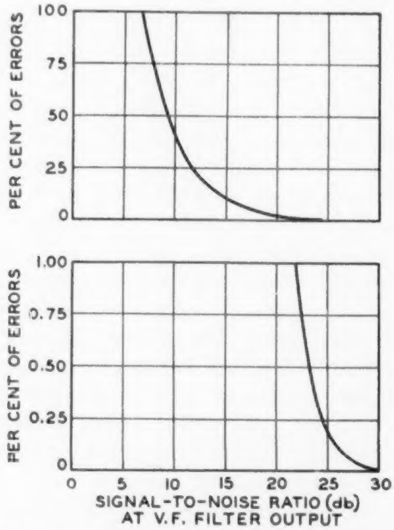


Fig. 5—Relation between printer errors and signal-to-noise ratio as determined from the Rocky Point-to-Rochester tests.

Since such a good correlation had been observed between the Rochester and Houlton data over the period covered by the tests, it appeared that the Rochester data might be extrapolated to cover a greater time by use of the Houlton noise readings. The same general method as outlined in Appendix A has, therefore, been applied to the Houlton loop noise data for the entire year of 1930 and the results are shown by Fig. 7. From this figure the great seasonal and diurnal variation in grade of transmission is at once apparent. It must be emphasized that

* For detail of method see Appendix A.

the per cent of errors corresponding with the average noise condition is a much more significant figure than the average per cent of errors. For example, in the Rochester tests Fig. 6 indicates that the per cent of error corresponding to average noise condition is 0.28 per cent while the observed average of the daily per cents of errors is 6.44 per cent. It is more useful to know that half of the time the copy will be better than

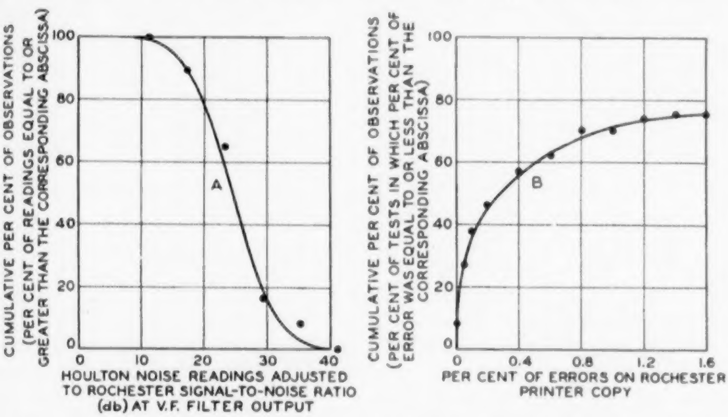


Fig. 6—Cumulative curves of signal-to-noise ratio derived from Houlton noise observations and of per cent of errors on the printer copy observed at Rochester.

0.28 per cent and half of the time worse, than to be unduly influenced by the effect on the average per cent of error of a few days in which the copy is almost all errors.

The results of the Rochester tests may be briefly summarized by giving a few figures which are based on the data obtained. A five-kilowatt station on long waves with a reasonable antenna, say 20 per cent efficient, would radiate one kilowatt. Assume that the local noise conditions are the same at the receiving station as those which have been used for the 9:00 P.M. values in Fig. 7 for Rochester, N. Y., variations. (These are obtained by applying a correction factor to the Houlton, Maine, noise observations for 1930.) Then the per cents of errors in the teletypewriter copy during the evening periods at different distances *

* As the distance varies between transmitter and receiver with the radiated power a constant, there is a variation in received signal field. If the noise is assumed to be fixed, this variation in distance will result in a variation in signal-to-noise ratio. Many of the commonly used radio transmission formulas take the form:[6]

$$E = \sqrt{P}\,\frac{300 \times 10^3}{D}\,\epsilon^{-\alpha D/\lambda^x}$$

For these calculations we have assumed $x = 1.25$ and from the field strength measurements at Rochester $\alpha = 0.023$. P is measured in kilowatts radiated, D and λ in kilometers, and E in microvolts per meter.

The various signal-to-noise ratios can then be translated into rates of error by use of Fig. 5.

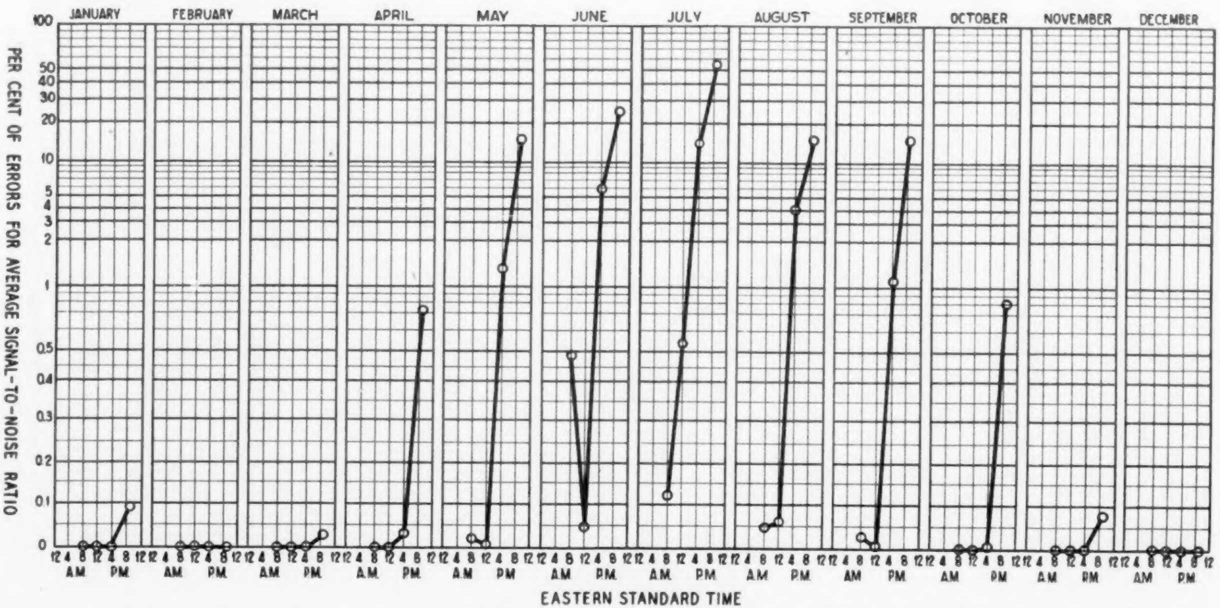


Fig. 7—Diurnal and seasonal variation of errors in printer copy to be expected over the Rocky Point-to-Rochester test radio circuit, based on Houlton noise observations in 1930.

from the transmitting station would be more than those given in the table for half of the time in each month.

TABLE II

Distance Overland from Station Radiating 1 kw at 60 kc (Statute Miles)	Errors in Printer Copy for Average Evening Noise Conditions for Each Month, Assuming Local Noise Conditions the Same as at Rochester, N. Y., and a Loop Receiving Antenna (Per Cent)											
	Jan.	Feb.	Mar	Apr.	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.
50	0	0	0	0	0	0	0	0	0	0	0	0
100	0	0	0	0	0.01	0.03	0.15	0.01	0.01	0	0	0
200	0	0	0	0.03	1.60	3.6	9.5	1.7	1.7	0.04	0	0
400	3.5	0.23	1.10	13.0	100	100	100	100	100	14.3	3.1	0.29

From these figures it is apparent that, under the conditions given, satisfactory all-year-round transmission could probably not be obtained over a radius of more than a hundred miles. To obtain the same grade of copy at a distance of 400 miles, as this assumed set-up could give at 100 miles, would require an increase in radiated power of about 25 db, making about 316 kilowatts radiated.

Development of systems and tests of the kind involved in obtaining information such as the authors have reported above have required the coöperative effort of a considerable number of engineers of the British General Post Office and of various parts of the Bell System. In solving many of the problems of telegraph signal transmission Mr. J. Herman was particularly active.

Appendix A

In deriving Curve A on Fig. 6 between "cumulative per cent of observations" and "Rochester signal-to-noise ratio at the voice-frequency filter output" from the Houlton noise data, the following facts were assembled and coördinated. In the first place, it was determined from the analysis of a large number of observations of loop noise at Houlton, that the magnitude of the noise is random and that its distribution obeys the Normal Law of Probability frequently used in engineering studies, provided the values of noise are in each case expressed as the number of decibels the "warbler" noise is above one microvolt per meter. Since each observation requires about the same time to complete and the observations are made at the same fixed times each day, the process really becomes one of sampling and the "per cent of observations" is equivalent to the "per cent of time" for the period covered by the tests. Then if, as in the Rochester tests, the radio signal strength is substantially constant, the signal-to-noise ratio (expressed

in db) becomes simply a constant minus the noise value (also expressed in db); and finally to get the signal-to-noise ratio at the voice-frequency filter output, a constant correction factor must be subtracted to take care of the band width, the difference in the methods used to measure noise and the difference in the absolute value of the noise observed at the two stations. Of course, if the Houlton loop noise is equal to or less than a given value for say 90 per cent of the time, the signal-to-noise ratio at the voice-frequency filter output derived from the Houlton noise will be equal to or less than its value for 10 per cent of the time.

Curve *B* of Fig. 6 is obtained directly from the observed errors on each test at Rochester and indicates in what per cent of the tests the per cent of errors observed was equal to or less than the value of "per cent of errors" given by the corresponding abscissa.

To combine the two curves of Fig. 6 it must be assumed that for each value of signal-to-noise ratio at the voice-frequency filter output there can be but one value for the observed per cent of errors, i.e., the variation in the per cent of errors depends only upon the signal-to-noise ratio received. If this is true, it is evident that a certain signal-to-noise ratio occurring a definite per cent of the time will always correspond to the per cent of errors which occurs the same per cent of the time. Hence, from the cumulative curves of Fig. 6 a curve relating signal-to-noise ratio with per cent of errors can be derived which is the same as Fig. 5. To do this a certain signal-to-noise ratio for which the corresponding per cent of errors on the Rochester printer copy is desired is selected. Curve *A* of Fig. 6 shows that this or some larger value of signal-to-noise ratio occurs *P* per cent of the time, but *P* per cent of the time, according to Curve *B* of Fig. 6, the per cent of errors on the Rochester printer copy was equal to or less than *E*. It is apparent, therefore, that *E* must be the value desired.

Assuming some constant received field strength at Rochester it is possible by this method to convert any individual Houlton loop noise observation into the corresponding per cent of errors on the Rochester teletypewriter copy.

Bibliography

1. Paul M. Rainey, "A new printing telegraph system," *Elec. World*, **65**, 848; April 3, 1915.
John H. Bell, "Printing telegraph systems," *Trans. A.I.E.E.*, **39**, 167, 1920.
H. P. Clausen, "Morkrum-Kleinschmidt printing telegraph systems," *Elec. Communication*, **5**, 216; January, 1927.
2. B. P. Hamilton, H. Nyquist, M. B. Long, and W. A. Phelps, "Voice-frequency carrier telegraph system for cables," *Jour. A.I.E.E.*, **44**, 213; March, 1925. *Trans. A.I.E.E.*, **44**, 327; 1925. *Elec. Communication*, **3**, 288; April, 1925.
E. H. Colpitts and O. B. Blackwell, "Carrier current telephony and telegraphy," *Trans. A.I.E.E.*, **40**, 205; 1921.

3. Fritz Van der Wonde and Alfred E. Seelig, "The high power Telefunken radio station at Sayville, Long Island," *Proc. I.R.E.*, **1**, 23; July, 1913.
4. Julius Weinberger, "The recording of high speed signals in radio telegraphy," *Proc. I.R.E.*, **10**, 176; June, 1922.
5. R. A. Heising, "Printing telegraph by radio," *Jour. Frank. Inst.*, **193**, 97; January, 1922.
   Henri Abraham and René Planiol, "La radiotelegraphie imprimee—emploi du telegraphe baudot en telegraphie sans fil," *Annal. des Postes, Telegraphes et Telephones*, **10**, 193; 1921.
   "The Baudot system and radio," *Electrician*, **93**, 207; August 22, 1924.
   Fritz Banneitz, "Über Versuche und Erfahrungen mit Drahtloser Schnelltelegraphic bei der Reichspost," *Elektrotech. Zeits.*, **42**, 714; July 7, 1921.
6. Lloyd Espenschied, C. N. Anderson, and Austin Bailey, "Transatlantic radio telephone transmission," *Elec. Communication*, **4**, 7; July, 1925. *Proc. I.R.E.*, **14**, 7; February, 1926.
   Ralph Bown, Carl R. Englund, and H. T. Friis, "Radio transmission measurements," *Proc. I.R.E.*, **11**, 115; April, 1923.
7. H. Nyquist, R. B. Shanck, and S. I. Cory, "Measurement of telegraph transmission," *Jour. A.I.E.E.*, **46**, 231; March, 1927.
8. John R. Carson, "Selective circuits and static interference," *Bell Sys. Tech. Jour.*, **4**, 265; April, 1925.
9. The Radio Corporation of America successfully tested printing telegraphy via their Chatham, Mass., radio station to the R.M.S. Majestic; *New York Times*, November 14, 1922.
10. The Navy Department successfully tested printing telegraphy between an airplane and ground radio station; *New York Tribune*, August 10, 1922; *New York Times*, August 10, 1922.
11. G. S. Vernam, "Cipher printing telegraph systems for secret wire and radio telegraphic communications," *Trans. A.I.E.E.*, **45**, 295; 1926.
12. S. B. Wright and H. C. Silent, "The New York-London telephone circuit," *Bell Sys. Tech. Jour.*, **6**, 736; October, 1927.
13. R. A. Heising, "Production of single sideband for transatlantic radiotelephony," *Proc. I.R.E.*, **13**, 291; June, 1925.
14. H. D. Arnold and Lloyd Espenschied, "Transatlantic radiotelephony," *Jour. A.I.E.E.*, **42**, 815; August, 1923.
15. Austin Bailey, S. W. Dean, and W. T. Wintringham, "The receiving system for long-wave transatlantic radiotelephone," *Proc. I.R.E.*, **16**, 1645; December, 1928.

Audible Frequency Ranges of Music, Speech and Noise *

By W. B. SNOW

This paper describes the use of an electro-acoustic system, transmitting the audible frequency range almost uniformly, in determining by ear the frequency ranges required for faithful reproduction of music, speech, and certain noises.

Sounds were reproduced alternately with and without filters limiting the frequency range transmitted by the electrical circuit. The filter cut-offs producing just noticeable changes in the reproduction were deduced from judgments of listeners as to the presence or absence of filters. It was found that for absolute fidelity all musical instruments except the piano require reproduction of the lowest fundamentals. The frequencies above 5000 cycles were shown to be important, some instruments and particularly noises requiring reproduction to the upper audible limit.

Tests were made in which experienced listeners judged the degradation of "quality" produced by a series of filters. The judgments showed definitely that the quality continues to improve as the frequency range is extended down to 80 or up to 8000 cycles. Although somewhat indefinite on cut-offs outside these limits, they indicated that reproduction of the full audible range was considered most nearly perfect.

ANY sound transmission system, if it is to give faithful reproduction, should transmit all the audible frequencies of a sound in their proper relative intensities. To give acceptable reproduction, it should transmit those frequencies considered most necessary for any particular application. The audible frequency range depends upon physical factors—the frequency-amplitude characteristics of a sound and the hearing characteristics of the average ear—whereas the acceptable frequency range must be determined by judgment when engineering or economic considerations limit transmission. As engineering limitations disappear and practical design becomes more a matter of economics a knowledge of both audible and acceptable limits increases in importance.

The program of listening tests described in this paper was undertaken primarily to establish the audible frequency ranges of the sounds most often encountered in sound reproduction, but some tests bearing on acceptable ranges were included. The sounds were transmitted through an electro-acoustic system equipped with electrical filters by means of which all frequencies above or below any desired cut-off could be suppressed, and observers determined the high and low frequency cut-offs causing just perceptible differences in the transmission. All

* Presented at the Camden mtg. of the Acous. Soc. Amer., May 4–5, 1931. Published in the Jour. Acous. Soc. Amer., July, 1931.

audible frequencies of the sounds were included in the range between the cut-offs thus delineated. Sound sources were: Musical instruments—tympani, bass drum, snare drum, 14″ cymbals, bass viol, 'cello, piano, violin, bass tuba, trombone, French horn, trumpet, bass saxophone, bassoon, bass clarinet, clarinet, oboe, soprano saxophone, flute, and piccolo; male and female speech; noises—footsteps, hand clapping, key jingling. These tests are described in Part I.

Measurements of the relation of reproduced frequency range to the quality of orchestral music, as judged by a number of experienced listeners, are reported in Part II. Tests of this kind must be used in establishing acceptable frequency ranges.

Part I

Apparatus

The reproducing equipment was built by the Bell Telephone Laboratories especially for fundamental studies of speech and music quality. Fig. 1 is a block diagram of the circuits involved in these tests. The

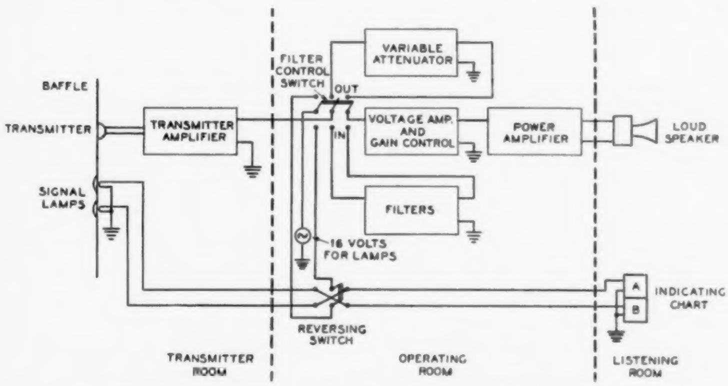


Fig. 1—Schematic circuit diagram of essential apparatus.

electro-dynamic microphone was mounted in a 5′ square baffle placed near the center of a large soundproof room, 29′ × 29′ × 13′ in size, which had a reverberation time of about one second for frequencies between 60 and 4000 cycles. The microphone amplifier, mounted at the rear of the baffle, raised the microphone output to a level that permitted satisfactory switching operations without objectionable surges.

In another room the filters, their switching circuits, and the main amplifiers were set up. The attenuator shown in the "filter-out" circuit was used to compensate for the losses in the transmitted bands of

the filters, so that the passed frequencies were reproduced at constant level at all times. Filters available were: high pass 30, 40, 55, 75, 100, 125, 250, 375, 500, 750, 1000, 1500 cycles cut-off frequency; low pass 13,000, 10,500 8500, 7000, 5500, 4500, 3750, 3250, 2850, 2450, 1900, 1500, 1000 and 750 cycles cut-off frequency. All were composite structures giving sharp cut-offs and attenuations of 60 db or more in the attenuated region. Representative attenuation characteristics are shown in Fig. 2.

The loud speaker was mounted in one corner of a third room, of dimensions 18′ × 27′ × 15′, semi-sound proof in construction and exhibiting reverberation characteristics similar to those of the microphone room. To cover the required frequency range, two reproducing

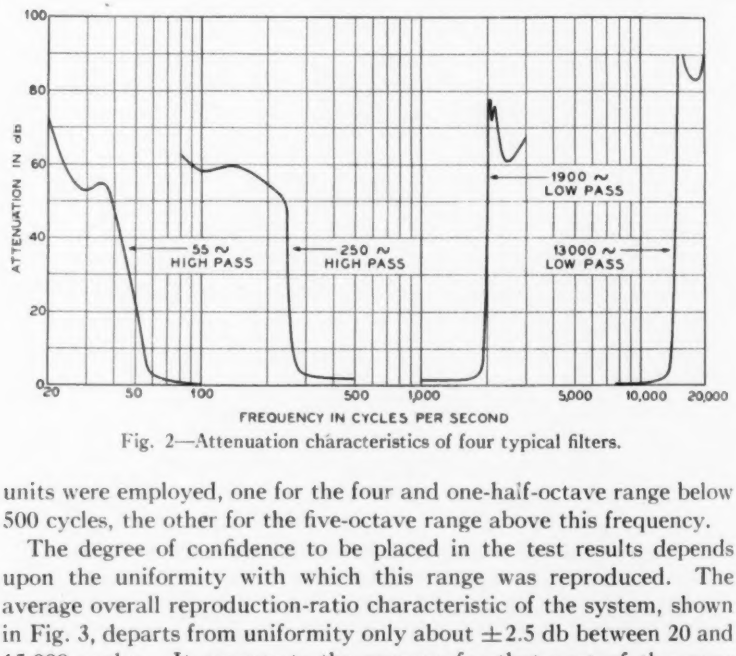


Fig. 2—Attenuation characteristics of four typical filters.

units were employed, one for the four and one-half-octave range below 500 cycles, the other for the five-octave range above this frequency.

The degree of confidence to be placed in the test results depends upon the uniformity with which this range was reproduced. The average overall reproduction-ratio characteristic of the system, shown in Fig. 3, departs from uniformity only about ±2.5 db between 20 and 15,000 cycles. It represents the average for that part of the room which may be called the "listening area," the directional characteristics of the loud speaker not permitting uniform sound pressure throughout the room at very high frequencies. At no point in this area did the measured pressure at any frequency depart more than ±3.5 db from the average curve. One assumption is involved. Because the measurements were made by supplying "warbling" frequencies to the volt-

age amplifier and measuring the sound pressure in the listening room with the regular system microphone and amplifier, it is necessary to assume that the microphone behaved identically in the microphone room. The two rooms are similar and the assumption was thought justified. The power output capacity of the system was estimated at one-half watt peak sound power with 10 per cent distortion products.

In addition to the speech circuits, Fig. 1 shows an indicating-lamp circuit. Placed before the loud speaker was a small box bearing the letters A and B on its translucent face. As the filters were thrown in or out the illumination was changed from one letter to the other, the letter corresponding to "filter-in" being determined by the reversing switch. The signal lights beside the microphone were lighted whenever the circuit was closed through. An order-wire circuit (not shown) was used for signalling and intercommunication.

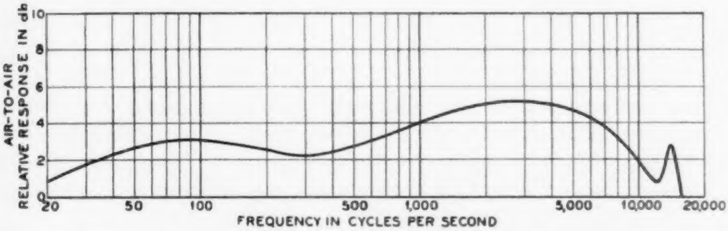


Fig. 3—Reproduction-ratio characteristic of complete system.

Testing Methods

The "A-B Test" method was used in determining the filter cut-offs producing just perceptible changes in the sounds. The observers listened to two conditions, A and B, one filtered and one unfiltered, and judged which condition was filtered. If the observer obtained a score of 100 per cent correct judgments in a large number of trials the filter was absolutely detectable. A score of 50 per cent correct judgments indicated an undetectable filter, because the observer, if guessing, guessed right and wrong an equal number of times.

Nine members of an articulation testing crew and two young engineers made up the regular observing personnel, though other observers were secured when possible. The actual number participating in the tests varied from nine to fourteen. All were known to have normal hearing, but the predominantly youthful makeup of the crew probably made the crew's average sensitivity for very high frequencies somewhat greater than the general average. The observers were frequently shifted about to insure average results, since the sound field was not absolutely uniform.

Professional musicians were employed in all tests with musical instruments. Usually they were seated as close to the microphone as practicable, and the amplifiers were set so that the sounds were reproduced at natural loudness. The power capacity of the system did not permit this loudness on the drums, cymbals, piano, trumpet, and trombone. For these cases the performers were seated about 10 feet from the microphone and the amplifier gain was reduced the necessary amount. Speakers were seated with their lips 18 inches from the microphone. The keys were shaken about four feet away. Hand-clapping and footsteps were produced at a distance of 15 feet.

The musicians were instructed to play their instruments "loud," as listening tests showed that the widest frequency ranges were thereby produced. Tests were made with the instruments played in their several octave ranges or with their different techniques to insure "boundary" results. In general the performers played repeated three or four note scales, for the differences produced by the boundary filters were too small to be detected regularly except on repeated music not supplied by melodies. However, such a procedure would not be representative for the piano, and regular player rolls were used in testing it. One repeated over and over a 15 second passage emphasizing the notes of fundamentals 32 to 800 cycles, the second similarly emphasized notes of fundamentals 200 to 3500 cycles, while the third was a march covering the range 40 to 1500 cycles.

Before the regular crew started work on each sound the engineers in charge listened to the reproduction and picked out the playing techniques that promised the widest frequency ranges. Tests always started with a filter giving 90–100 per cent correct judgments and continued through successive cut-offs until the 50 per cent cut-off was reached. Throughout both the preliminary and regular tests the observers made notes relative to quality changes produced by the filters, and noises produced by the instruments. With the performers and observers in readiness for an actual test the procedure was as follows: The filter operator threw the main switch from neutral to the position lighting the "A" lamp, which might be "filter-in" or "filter-out" as he chose. The performer, seeing his signal lamp light, then played his instrument for a period of 15 to 20 seconds as the operator switched "A-B-A-B-neutral." Switching to neutral stopped the musician by extinguishing his signal light, and gave the observers an opportunity to check on their recording blanks the condition they believed to be filtered. The process was repeated five times with a random order of correspondence between A or B and "filter-in." When necessary a filter was retested until practice effects were eliminated. Since there

were never less than nine observers, and each had at least six trials on each filter, the minimum number of observations used in computing the percentage of correct judgments on any filter cut-off was 54. Several times check tests were made in which the lights were changed, but no filter was inserted. The average scores on these tests always were within the limit 50 ± 4 per cent.

Data

The filter cut-offs producing just noticeable effects upon the sounds were not sharply defined. For every sound there was, between the cut-off recognized every time and the cut-off never recognized, a certain region of appreciable width where the percentage of correct judgments decreased from 100 per cent to 50 per cent. If this percentage is plotted against cut-off frequency a curve such as is shown in Fig. 4

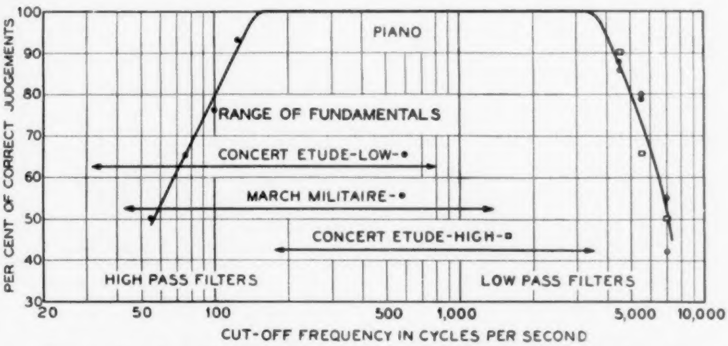


Fig. 4—Number of times filter condition was correctly perceived as function of cut-off frequency—piano.

results. Curves of this kind proved useful for interpolation purposes, but their contours were felt to be too dependent upon the individual peculiarities of observers and players to be of general significance. No close correlation existed between the importance of any frequency range and the contours of the curves, for the differences caused by the filters that were recognizable in less than 80 per cent of the tests were very small. In addition, some observers consider that elimination of high frequencies improves the reproduction of certain musical sounds by removing accompanying noises. Therefore it was decided that the useful information from the data could best be presented by straight lines.

The audible frequency ranges of all the sounds tested have been plotted in this way in Fig. 5. The end points for these lines have been

taken where the correct judgments amounted to 60 per cent. In addition, the frequencies where 80 per cent correct judgments were obtained have been marked. The region between the 80 per cent marks is the most significant—frequencies above and below would probably not be heard at auditorium distances or with other instruments playing. There were certain sounds that apparently extended to the highest audible frequencies as judged by the ease with which the highest filter—

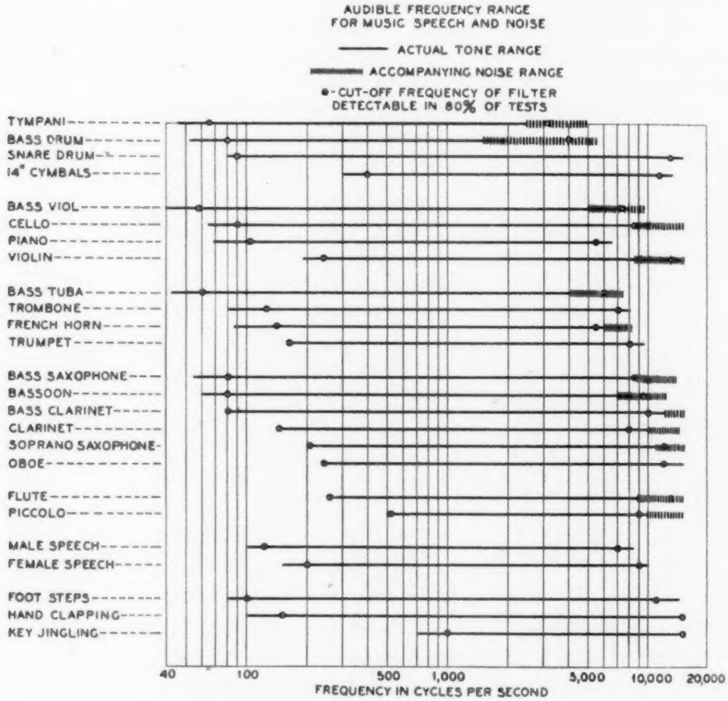


Fig. 5—Audible frequency ranges.

13,000 cycles cut-off—was detected. In these cases the lines have been arbitrarily stopped at 15,000 cycles.

With most instruments it was noticed that the actual musical sounds were accompanied by other sounds such as key clicks, lip noises, "buzz" of reeds, and hissing of air. The observers tried to distinguish between the frequency ranges carrying the two classes of sounds, and their judgments are summarized by the treatment of the lines of Fig. 5. The heavy portions indicate the frequency ranges thought to convey the "tone quality" of the instruments, and the short vertical lines

define the ranges of noise. In some cases noise and tone seemed inseparably blended.

The qualitative observations made by the observers are summarized in the notes below, in which "L. F." means "lowest fundamentals."

Tympani—No important frequencies below 65 cycles (drum tuned to 96 cycles). Actual tone range ends around 2000 cycles. Prominent drum rattle and beating noises to around 5000 cycles.

Bass Drum—No important frequencies below 70 cycles. Actual tone range ends around 1000 cycles. Prominent drum rattle and beating noises to around 5000 cycles.

Snare Drum—No important frequencies below 100 cycles. Actual tone consists of rattle extending to very high frequencies.

14" Cymbals—No important frequencies below 350 cycles. Low frequencies prominent when one cymbal is struck with a hard stick. High frequencies prominent when two cymbals are clashed together.

Bass Viol—L. F. fairly important, slightly more on plucked than on bowed notes. Considerable bowing noise.

Cello—L. F. fairly important. Tone very rich in harmonics. Moderate bowing noise.

Piano—L. F. unimportant for first octave. 100 cycle high pass filter only slightly noticeable. Upper notes practically pure tones.

Violin—L. F. important. Tone rich in harmonics. Noises and tone blended.

Bass Tuba—L. F. fairly important. "Pedal" notes—fundamentals around 20 cycles—contain fewer very low frequencies than regular notes. Moderate blowing and key noises.

Trombone—L. F. not very important below 130 cycles. Middle register has greatest harmonic content. Inappreciable noise.

French Horn—L. F. unimportant below 130 cycles. Middle register has most volume and harmonics. High register gives rather pure tones. Harmonics least prominent of any instrument tested.

Trumpet—L. F. fairly important. Lowest register has greatest high frequency "blatt." Tones purer at higher pitches. Inappreciable noise.

Bass Saxophone—L. F. not very important below 90 cycles. Highest register rather unmusical and unpleasant. Considerable blowing and key noise.

Bassoon—L. F. fairly important. Prominent reed noise on lower register. Moderate key slap.

Bass Clarinet—L. F. very important. Tone goes to very high frequencies on upper register. Prominent reed noise on lower register becoming blended with tone on upper register.

Clarinet—L. F. very important. Medium range has largest harmonic content. Highest range gives much purer tones. Moderate blowing and reed noises at very high frequencies.

Soprano Saxophone—L. F. very important. Powerful harmonics making very harsh tone. Moderate reed noise above 10,000 cycles, less than that of clarinets.

Oboe—L. F. important. Most "reedy" tone of all tested. Tone extremely rich in harmonics of high order, especially middle register. Noises blended with tone.

Flute—L. F. very important. Middle register has most harmonics. Highest register produces almost pure tones. Much blowing and mechanism noise on highest register.

Piccolo—L. F. very important. Middle range most musical and free from noise. Highest few notes are very powerful but are practically pure tones. Much blowing noise and rumble on all registers.

Footsteps—No important frequencies below 100 cycles. High frequencies up to about 10,000 or 12,000 cycles required.

Handclapping—No important frequencies below 150 cycles, but requires the entire audible range on the high frequency end. Sounds fairly natural with 8500 cycle cut-off.

Key jingling—bunch of 22 keys shaken on 4" wire loop—No important frequencies below 500 cycles but requires entire audible range on the high frequency end. Tone very unnatural with 8500 cycle cut-off.

It is felt that the caliber of the playing was such as to render the comments and measured frequency ranges generally applicable. These ranges probably represent extreme conditions, for the observers were in effect situated unusually close to the instruments, they were listening under most favorable conditions, and they had only to pick out a particular distortion.

The piano was the only instrument which did not require the reproduction of its lowest fundamentals for perfect fidelity. Therefore transmission of 40 cycles—the lowest note of the bass viol—was required, and this was found to be ample for the percussion instruments. However, as the 80 per cent marks indicate, little was lost when frequencies below 60 cycles were not reproduced.

Many of the instruments produced noises that extended to high frequencies, but only the oboe, violin, and snare drum were thought to extend their tone ranges to the upper audible limit. The action of bows on the strings and the clatter of reeds in the reed instruments produced very prominent noises of high frequency. When the lips were used as reeds the noises were much less prominent. The noises indicated for the flute and piccolo were produced by the impact of the air from the lips against the embouchure opening. As a group the lipped instruments produced only moderately high frequencies; the other groups all had some instruments producing frequencies extending to the upper audible limit. An upper cut-off of 10,000 cycles did not affect the tone of most of the instruments to a marked extent, but every instrument except the bass drum and tympani was affected by the 5000 cycle cut-off. A frequency range of 100 to 10,000 cycles was shown to be entirely satisfactory for speech.

Between the 80 per cent marks the bass viol required the greatest range—7 octaves—and the piccolo required the smallest range—4 octaves.

Noises in particular were characterized by high frequencies. Hand clapping and key jingling were both found to be very definitely changed by the 13,000 cycle filter, and informal listening tests on several other noises indicated that high frequencies were very prominent. Probably many noises also contain important frequencies below 100 cycles and transmission of the entire audible range would seem much more important for noise reproduction than for reproduction of musical sounds.

Part II

The measurements of the quality changes produced by the filters were made using the same apparatus but a different testing technique. The 18 piece orchestra furnishing the music was made up as follows:

3 first violins, 1 second violin, 1 viola, 1 'cello, 1 string bass, 1 flute, 1 oboe, 2 clarinets, 1 bassoon, 2 French horns, 2 trumpets, 1 trombone, 1 drummer. The players were seated in concert arrangement with the violins about 8′ from the microphone. Ten engineers experienced in quality judgments acted as observers. In these tests the filter conditions were always presented as "B" and the observers were asked to rate the quality of the "B" condition numerically, considering the "A" condition to possess a quality of 1.0. The ratings could be either less than 1.0, indicating a degradation, or greater than 1.0, indicating an improvement. Conditions were switched A-B-A-B—, continuing until all observers had obtained a judgment, but the filters were presented in

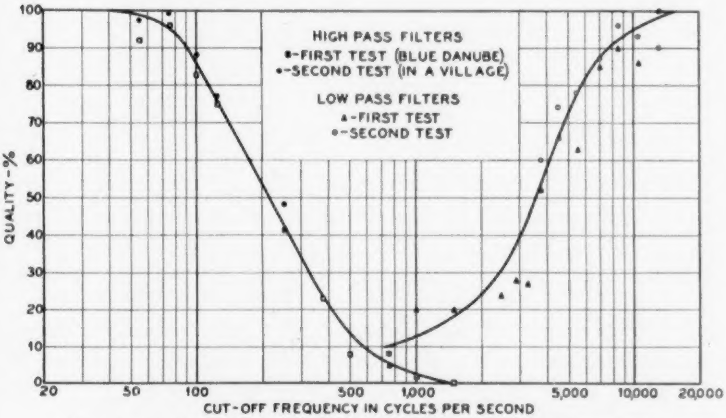


Fig. 6—Quality of orchestral music as function of cut-off frequency.

perfectly random order. Therefore the observers were never informed as to what filter was being tested; they only knew that "A" represented a quality of 1.0 and "B" a condition to judge.

The orchestra played the Strauss waltz, "The Beautiful Blue Danube," for the first test. It was orchestrated so that most of the instruments played most of the time. All filters except 30 and 40 high pass were presented to the observers. From the results a list of filters which all observers had rated as better than 0.5 was compiled for presentation during a second test. This time the music was "In The Village," a composition of Godard. It was a selection in which many instruments had solo parts and in which therefore the character of the music changed rapidly.

The average ratings for both runs are plotted in Fig. 6. The two sets of data agree reasonably well except at the extreme ends where

differences are small and the judgments are greatly affected by particular instruments. Clearly the quality rises rapidly as the cut-off is extended upwards to 8000 cycles, or downward to 80 cycles, but outside these limits the results are inconclusive. Out of 370 ratings recorded, only a scattered 13 were greater than 1.0. In general, therefore, it must be concluded that reproduction of the full audible range was preferred.

The curves of Fig. 6 do not define acceptable frequency ranges directly, but the method with slight changes would give them. The observers would be instructed to judge whether for any particular application the range being heard would be satisfactory. However, these curves, coupled with the general experience of engineers and musicians should aid in determining acceptable ranges where direct tests are impracticable.

Conclusion

The author is not familiar with any published results of comprehensive listening tests that can be compared directly to these data. However, the audible ranges here presented have been compared with physical measurements [1] of peak sound output of a number of the instruments. The physical measurements give the peak amplitudes in octave ranges below 500 cycles, and in half octave ranges above this point, whereas interpolation between these limits was possible in selecting the audible ranges. On the other hand, auditory masking must play a part in determining the audible cut-off points. Considering these limitations to comparison, the two sets of data are consistent on every instrument tested in common.

The more important results of the tests are considered to be as follows:

1. The piano was alone in producing tones with inaudible fundamentals.

2. Audible frequencies down to 40 cycles were produced by the musical instruments, but reproduction only to 60 cycles was considered almost as satisfactory.

3. It was found that transmission of the highest audible frequencies was needed for perfect reproduction of musical instruments, mainly because of the noises accompanying the musical tones. A 10,000 cycle upper cut-off had slight effect upon the tone quality of most instruments, but a 5000 cycle cut-off had an appreciable effect upon all except the large drums.

[1] "Absolute Amplitudes and Spectra of Certain Musical Instruments and Orchestras" by L. J. Sivian, H. K. Dunn and S. D. White, *Jour. Acous. Soc. of America*, January, 1931.

4. The quality of reproduction of orchestral music continued to improve materially as the lower cut-off was extended to about 80 cycles and the upper cut-off to about 8000 cycles. Reproduction of the full audible range was preferred to any limitation of band width.

5. Noises required reproduction of the highest audible frequencies. A 10,000 cycle cut-off caused appreciable reduction of naturalness on common noises. It was felt that this cut-off probably would never preclude recognition of a noise.

The results seem to necessitate no radical revision of the qualitative ideas entertained by many acoustical engineers for some years, their value lying rather in the quantitative corroboration they supply.

Contemporary Advances in Physics, XXII

Transmutation

By KARL K. DARROW

In this paper are described experiments made at the Cavendish Laboratory, in Vienna, in Chicago and elsewhere during the last fifteen years in which atom nuclei have been disrupted by swiftly moving alpha particles ejected by radioactive materials. Whether an atom is an atom of gold, or of tin, or of praesodymium, or one of some other of the ninety-two varieties is determined solely by the magnitude of its nuclear charge.

When the disruption of atom nuclei occurs spontaneously, as it does among atoms of the radioactive elements, the fragments are nuclei of charge different from that of the exploded atom. When disruption is brought about by design, as it is in the experiments described in this article, we have again the disappearance of atoms of one species and the appearance of atoms of other species. The cases differ in that in the first the action goes on without let or hindrance, while in the second it is, to a certain extent, under the control of the experimenter. The experimenter may, if he chooses, congratulate himself on having solved the age old problem of the transmutation of elements. However, transmutation as such is not the object of these investigations. If it were their success would have to be rated as altogether negligible, for the quantities of material transmuted are much less than can be detected chemically.

The real object of the work, as is made abundantly clear, has been to verify and to extend our knowledge of the constitution of the nucleus. This is a subject about which a great deal is yet to be learned, but one on which the physicist has already many strong convictions based on a considerable array of interrelated and consistent data. The various conclusions regarding the constitution of atom nuclei which had been reached before any of these experiments on artificial disintegration were made, are discussed. The results of the investigations described confirm and extend our knowledge of the constitution of the nucleus.

IT is often said that the conversion of the elements into each other has been the dream of the human race for many centuries, nay even for millennia. In special and practical cases, this is probably true; I suppose that from the dawn of history most men possessed of stores of lead or silver have tantalized themselves by dreaming of these being changed to gold. But in the general sense it must be false. One cannot aspire to transform element into element, if one does not know what elements are; and no one had such knowledge centuries ago. Surely many of the chemical reactions which we consider commonplace, many of the compounds old and new which the modern chemist makes in his routine, would have seemed to ancient or to mediæval no less wonderful than any "transmutation" which he could possibly imagine. Could the Florentine or the Greek have been much more amazed by a change of silver into copper, than by the synthesis of a dye out of tar or coal, the growth of a diamond out of black carbon in a furnace?

It seems unlikely; for the very special wonderment and admiration, which the first-named change would evoke from a scientist of today were it achieved before him, would arise out of a wisdom denied alike to Greek and Florentine. Only the modern can know how much it would differ, in what particular way it would transcend all that has gone before.

For his power of appreciation, this modern onlooker would have two sciences to thank. These are the chemistry of the nineteenth and the eighteenth century, with its uncounted and uncountable attempts to analyze and synthetize and convert and transform, which led to the eventual conclusion that underneath the endless and changeable variety of visible matter there are certain substances which can neither be synthetized nor analyzed nor converted one into another; and the physics of the twentieth, which penetrated deep into the atoms of these unalterable substances, and there discovered the recondite and all-but-unassailable part, in which the character of each element is conserved. The former proved that all known chemical changes are made by combining these atoms or tearing them apart; the latter showed that what happens in every such event may be imagined as a rearrangement of flocks of electrons, which form the outer part of every atom. But chemistry further proved that such a change as that of silver to copper, or of mercury to gold, must of necessity involve something far more radical—something which the physics of alpha-particles and X-rays eventually made clear: in the atom there is an innermost *nucleus*, the centre of attraction whereby the electron-flocks are held together: this it is which must be reached and altered, if one element is to be transformed into another.

This statement, being as it is a description of the geography of the atom—perhaps I should say, a description of its astronomy, for these ultimate particles of matter are to be likened to a solar system rather than the earth—requires to be proved by exploration. The explorers sent out for this purpose are alpha-particles—corpuscles which are recognizable as such, for if they strike against a fluorescent screen each makes its separate luminous splash. Less than 10^{-12} of a centimeter in diameter, they are small enough to penetrate the electron-flocks of the atoms, which are spread over spaces tens of thousands of times as wide. The electrons near which they pass deviate them but little, being of less than one seven-thousandth their mass. Endowed with energy which may be as great as an electron could acquire from a potential-rise of *eight million* volts, they are able to approach the positive portions of the atomic structure though positively charged themselves. They are, in fact, extremely well fitted for the task of

exploration which in 1911 Rutherford imposed upon them, and of which they reported to him that in the atom there is a massive particle positively charged, like themselves less than 10^{-12} centimeter in diameter. They could not perceive the electrons which surround this "nucleus," bearing charges of which the sum compensates its own; but other evidence makes us secure of the existence of this flock, and of the general theorem that *the atom of the Nth element of the periodic table consists of N electrons surrounding a nucleus of the tiny dimensions aforesaid, having a charge + Ne and a mass almost the same as the entire mass of the atom.*

This last, then, is the entity which anyone must attack who wishes to transmute the atom. It takes no part (as I remarked above) in chemical phenomena, in the emission of light or of X-rays, in the electrical effects which atoms can achieve when they lose charge and so become ions. This for the wouldbe transmuter is a fact of serious import; for if the nucleus has no influence on these, no more have they on it. Radioactivity, indeed, is a quality of nuclei—radioactivity *is* transmutation, natural and spontaneous; and it is not affected by anything chemical or electrical, by any temperature or any illumination which has ever been applied to a self-transmuting substance. These kernels of the atoms are well sheltered and highly resistent; they seem as oblivious of the world around them, as the interior of the earth is unconscious of the life upon its surface.

But the properties of the alpha-particle which enable it to penetrate to the neighborhood of the nucleus—extreme minuteness, high momentum, enormous store of concentrated energy—may they not also qualify it to impinge directly on the atom-kernel, to invade the nucleus and disrupt it, to shatter it if it be shatterable at all? We may be sure that the nuclei of atoms, hydrogen perhaps excepted, are complex. They cannot be the ultimate and irreducible particles of matter; for radioactivity proves that some of them disintegrate of themselves into smaller and lighter bits, while as for the rest, the facts that the charges of all are multiples of a common charge and the masses of all are nearly multiples of a common mass must surely be taken as meaning that all of them are structures built of electrons and protons. In principle, therefore, they must be breakable, if only they can be struck with sufficient force by hammers of suitable size. Now of all known vehicles of available force, alpha-particles best combine the qualities of smallness and great energy, and therefore seem the best adapted to the task.

Such must have been the ideas of Rutherford; it may be presumed that he was meditating them during the war, since in the first year thereafter he put them to the test, and so became the first to achieve

transmutation beyond the shadow of a doubt. There was of course no certainty beforehand that he would succeed. On the contrary there were apparently grave grounds for pessimism. We must take note of these; the proof that they were not justified is not the least important part of Rutherford's achievement.

First, even the energy of the alpha-particles might have been too small to injure a nucleus. Indeed, for many kinds of atom-kernels it *is* too small, to judge from the work of Rutherford's school at the Cavendish Laboratory; and for the rest there is not much margin to spare; from the work of that school it appears that if the fastest alpha-particles moved with a speed as great as six-tenths as their actual speed, and no greater, the effect would never have been discovered.

Second, there was reason to fear that the nuclei are too small to be struck except by the rarest of chances, too rare a chance to be serviceable. The observations on deflected alpha-rays had proved that the kernels of atoms are less than 10^{-12} cm. across, the impinging particles no greater: a very thin missile and a very tiny target! Had they been a few orders of magnitude smaller than this maximum limit, "square hits" would have been too few to notice. As a matter of fact, in the first of the successful experiments, the proportion of these was about one to every million of alpha-particles traversing the layer of nitrogen gas which Rutherford was trying to transmute.

Third, the fragments of the broken nuclei might not have been observable. Delicate as are the methods of chemical analysis, they are not fine enough to detect alterations so infrequent as these were expected to be, and were actually found to be. Alpha-particles themselves are detected in three ways—by the luminous splashes or "scintillations" which they cause when they impinge on fluorescent screens; by the trails of water-droplets which they leave behind them when they dash through moisture-saturated air which is suddenly cooled just before or just after their passage; and by the electrical discharges (small-scale sparks) which they touch off when they pass through air in the neighborhood of a charged and sharply-pointed needle. The two last of these are due to ions which the particle forms by detaching electrons from molecules of the gas. The slower the particles, the fewer the ions; the less conspicuous are these effects and the more likely to be missed. As for the scintillations we know but little of their mechanism, but we do know that the slower the particles, the fainter the flashes. Thus it is altogether reasonable to suppose that when nuclei are broken into fragments, the fragments may be moving too slowly to be noticed by any of these three procedures! (One might even suspect that the pieces of a fractured nucleus may not have

the power of forming ions or evoking scintillations, however fast they move; but this would be too pessimistic; there is every reason to suppose that they are charged corpuscles, therefore possessed of the same powers as alpha-rays.)

In all likelihood, many atom-kernels are disrupted and their fate goes unperceived, because the "products of disintegration" move too slowly; but sometimes these are fast enough to be detected in any of the three aforesaid ways, as we shall see. There is, however, yet another peril. Consider the alpha-particles which pass close to nuclei without disrupting them. They are deflected, but the nuclei themselves suffer a reaction which sets them into motion. If these belong to elements of atomic weight greater than 30, or let us say 40 to err on the safe side, their masses are so much larger than those of the alpha-corpuscle that the speed they acquire is negligible. But if they belong to one or another of the half-dozen lightest elements, they may acquire a speed so great that of themselves they can make ions in a gas or scintillations on a screen. If an alpha-particle, being itself a helium nucleus, flies straight against the kernel of a helium atom but does not fracture it, then obviously the struck nucleus must take up the entire speed of the striking corpuscle. If it is a carbon or an oxygen nucleus which is thus squarely struck, without being broken, its final speed must be one half or four tenths that of the alpha-particle. And if it is a hydrogen nucleus or proton which is the victim of a square and central impact, it must go off with no less than *sixteen-tenths* of the speed of the impinger. Incidentally, the latter is slowed down to compensate for the kinetic energy acquired by the kernel which it strikes.

The dangerous consequence is, that in a stratum of matter of low atomic weight which is bombarded by alpha-rays, there must be intact but rapidly-moving kernels which may be confused, which indeed one can hardly help confusing, with the expected products of disintegration. Moreover, even in a stratum of an element of higher weight, a metal film or a tube of gas, there may be hydrogen enough to provide so many low-mass targets for the alpha-rays, that the region is filled with fast-flying protons which are not tokens of disruption. In every case where corpuscles are observed which are thought to be parts of fractured nuclei, it must be proved that they are not of this kind, nor yet are scattered alpha-particles.

Now as an index of the initial speed of an alpha-particle, people generally take its "range." The trail of water-droplets which the particle leaves along its path through suddenly-cooled moist air comes to a sudden end (Figs. 10, 11); the length of the trail, measured to its end from the point where the particle entered the air, is its range in the

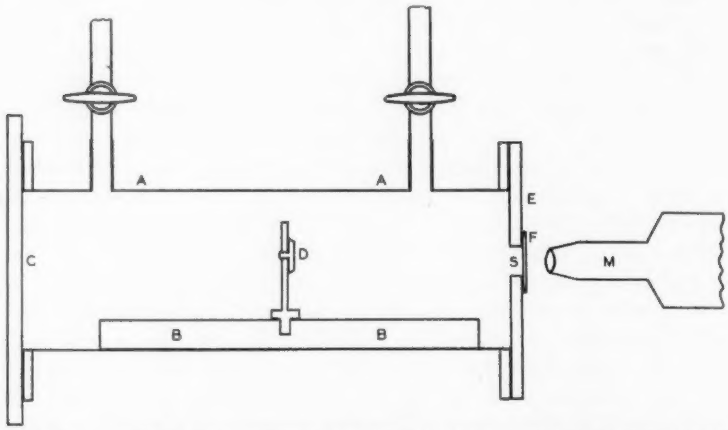
air, and depends in a known way (known by experiment) on the speed which the particle had at the moment of entry. Or the range may be determined by moving backward a fluorescent screen placed opposite the point of entry of the corpuscles; the scintillations are at first undiminished in number as the screen recedes, but eventually they cease, and cease quite suddenly; the distance to the point of their cessation is the range. In air at normal pressure and 15° temperature,[1] the range of the fastest known alpha-particles (apart from a few very scanty classes) fresh from the source is about 8.6 cm. It has been determined for a number of other gases as well, and for any gas it varies inversely as the density.

It follows then that if the fluorescent screen is placed at a distance from the source of alpha-particles so great that it lies beyond their range in the substance intervening, whatever scintillations may appear upon it are not due to alpha-rays. But it does not yet follow that the screen is beyond the reach of protons speeded up in the way I just described. To find out about this, it is necessary to know the relation between the speeds of protons and their ranges. Now the cause of the slowing-down and stopping of charged corpuscles, protons and alpha-particles alike, is this: as they flash through strata of matter, they tear electrons loose from the atoms which they pass, and spend their energy in doing so. The range of either sort of corpuscle is substantially the distance through which it can fly, before the major part of its initial energy is dissipated in this way. An alpha-particle has twice the charge of a proton, therefore extracts electrons oftener from the atoms near its course, therefore loses energy more quickly. If particles of the two kinds have equal range, the former must initially have had the greater energy. A theoretical analysis (achieved by Bohr and Darwin) shows that the ratio is that of the squares of the charges—four to one. But since the ratio of the masses is likewise four to one, the speeds are equal. Alpha-particles and protons of equal initial speed have (approximately) equal range. Now as I stated above, hydrogen nuclei struck centrally by alpha-particles acquire a speed 1.6 times as great as these, therefore, a range equal to that of alpha-particles moving 1.6 times as fast as those which made the impacts. It is a fact of experience that the range of alpha-particles varies about (not exactly) as the cube of their speed. If, therefore, hydrogen is bombarded by rays of a stated range R, hydrogen nuclei which suffer central impacts will be projected forward with ranges amounting to

[1] This is Rutherford's convention. Certain physicists specify the range in air at normal pressure and zero temperature, which stands to the other in the inverse ratio of the densities of the air, about 273 : 288. In later pages I shall occasionally adopt this usage.

$(1.6)^3R$, or about 4.1 times R. And if (for instance) air at normal temperature and pressure is bombarded by the alpha-particles of radium C' which in this gas have a range of seven centimeters, and scintillations are observed on a fluorescent screen beyond, the observer must reckon with the chance that they may be due to the nuclei of hydrogen molecules mixed with the air, so long as the distance to the screen is less than 4.1 times seven, or say thirty, centimeters.

On the principle that the best way to deal with a possible source of trouble is to examine it minutely, Rutherford prepared for his attempt at transmutation by a study of the nuclei which are struck and which



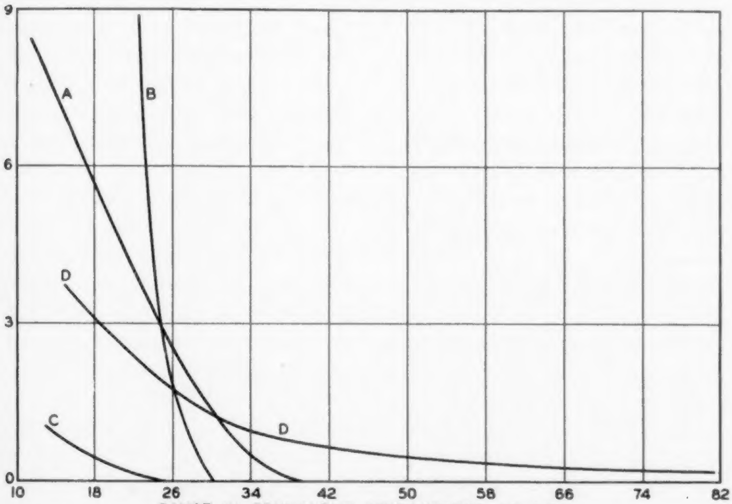
* Fig. 1—Rutherford's apparatus for detecting transmutation of gases by the scintillation method. Source of alpha-particles at D; gas in the tube; fluorescent screen (transparent) at S; microscope at M.

* From Sir Ernest Rutherford, James Chadwick and C. D. Ellis, "Radiations from Radioactive Substances," 1930. By permission of The Macmillan Company, publishers.

"recoil," as we say, when alpha-particles are fired into hydrogen. His pupil Marsden had begun on such a study in 1914, and had observed that scintillations appeared on a screen set up far beyond the ultimate reach of the alpha-rays—more than a hundred centimeters, inasmuch as in hydrogen the range of either kind of charged corpuscle is about four times as great as in air. Resuming the research in 1919 (one never needs to ask why things begun in 1914 should have lain so long uncontinued) Rutherford counted the scintillations and plotted their number for what, in effect, were various distances of the screen from the source. I must pause to say that in practice one does not draw the screen back so as to interpose thicker and thicker layers of gas between

and the point of entry of the alpha-particles; instead one leaves it fixed and varies the pressure of the gas, or else interposes a series of thin foils of aluminium or some other metal or of mica, each of which slows down the particles to the same extent (in the technical language, has the same "stopping-power") as a known thickness of air. (For instance, a thickness of mica of weight 1.43 mg. per square cm. is equivalent in stopping-power to 1 cm. of air at 15° C. and 760 mm. Hg.) In curves of the sort in which we shall be interested, number-of-scintillations is usually plotted along the vertical axis, number-of-centimeters-of-air along the horizontal; but in general some other substance did duty for the air, and its thicknesses were translated into equivalent thicknesses of this standard gas (at normal temperature and pressure) before the curve was drawn.

Curves of this sort appear in Fig. 2. All of them were obtained with gases bombarded by alpha-particles of seven-centimeter range. No-



* Fig. 2—Number of protons falling cn fluorescent screen, plotted as function of thickness of air which they have traversed since leaving the disrupted atoms.

* From Sir Ernest Rutherford, James Chadwick and C. D. Ellis, "Radiations from Radioactive Substances," 1930. By permission of The Macmillan Company, publishers.

tice first the curve marked *B*: it corresponds to hydrogen, mixed with carbon dioxide; and it testifies that the scintillations did not cease until the screen was shielded by the equivalent (in mica) of thirty centimeters of air, the amount computed for the range of hydrogen nuclei

struck centrally by alpha-corpuscles as fast as these. (That many of the nuclei causing scintillations did not have so great a range is easily accounted for; it is due to the fact that most of the impacts are sensibly "off-centre," the struck particles flying off obliquely with less energy than they would have derived from a "square hit.") But at thirty centimeters of "air-equivalent," they cease entirely; this sustains us assuming that if with any other gas or any solid there are scintillations when the screen is so much shielded, they cannot be due to admixtures of hydrogen.

Curve *C* was obtained with oxygen; what there is of it is ascribed to commingled hydrogen; in any case, it does not extend beyond the critical point at which, were there any flashes still to be seen, they could safely be attributed to something else.

Curve *A* is more sensational: very definitely it extends beyond the critical length; very definitely there are corpuscles able to make their way through deeper strata of matter than either the primary alpha-particles or such nuclei of stray hydrogen atoms (so both theory and experiment assure us) as these might find to strike. This curve was obtained with air. Since with pure oxygen there was no sign of such extraordinary corpuscles, it is to be presumed that they were due to the other of the major gases of the air—an inference which the study of pure nitrogen made sure.

The most astonishing of all the curves is *D*. It stretches far beyond the critical point; flashes appeared on the screen when even as much as the equivalent of ninety centimeters of air lay between it and the substance which the alpha-rays were striking, which was aluminium in the form of a thin leaf. Thus, when foil of aluminium is subject to the impacts of these rays, it throws out corpuscles three times as penetrative as the very fastest which a critic might possibly discredit by ascribing to occluded hydrogen.

Are these, then, fragments of disrupted nuclei of aluminium or nitrogen? and are they protons?

In principle the second question is answerable by itself; it is sufficient to deflect the corpuscles by electric and magnetic fields, and measure their deflections; the value of their charge-to-mass ratio (which if they are protons is about .00054 of the value for an electron) could then be computed, and incidentally their speed also, which itself would be well worth determining in a way more direct than by inference from the range. But though such measurements have many times been made on other kinds of particles, and the technique is very well developed, the application to those of this especial kind is difficult because they are so few. Say that the apparatus is so built that they

fall on only a part of the screen; if a magnetic field is applied in the proper sense to the region which they traverse, the spot on which they fall moves sidewise; but the flashes are so infrequent that the shift is not obvious, and only by lengthy countings can one be sure that more of them appear in one place and less in another when the field is on than when it is off. Rutherford however managed to make countings enough to prove that the shift is of the order of magnitude to be expected, if the particles are protons having the speed inferred from their range; and incidentally that they are positively charged, something which has been taken for granted but which requires proof. It was with the long-range corpuscles expelled from aluminium, from phosphorus and from fluorine that he achieved these results.

The problem was then taken up by Stetter in Vienna; he tried the scheme developed to so high a pitch by Aston in his famous series of experiments on isotopes—a scheme of which I shall say only that although it involves both electric and magnetic fields, they are so arranged that corpuscles having a common value of charge-to-mass ratio are brought to a common focus irrespective of their speeds (so long as these are not dispersed over too wide an interval); therefore, by locating the focus, one may recognize the kind of corpuscle. In Fig. 3 appears a part of his apparatus: the source of alpha-particles at Q, the sheet of transmutable substance at S or S', and beneath it the system of long narrow parallel channels which Stetter arranged so that only a beam of corpuscles following almost perfectly parallel paths should enter the deflecting fields below.

Shifting from place to place the microscope with which he examined the screen beyond the deflecting fields, and counting the scintillations, Stetter found three foci which in the curve of Fig. 4 appear as three peaks (the ordinate being the number of flashes in unit time over a given area, the abscissa the distance of the midpoint of this area from a point taken as zero). From the positions of these foci on the screen it followed that the one on the right was due to corpuscles having the charge-to-mass ratio of protons; the one in the middle, to alpha-particles; the one on the left of which only a part appears, to corpuscles having a charge-to-mass ratio half as great as that of an alpha-particle. Two then are proof of alpha-rays deflected by the metal at S, some of which had lost one-half of their positive charge through picking up an electron somewhere in their careers; the third is evidence of protons, and strong evidence, for Stetter estimates the uncertainty of his measurement of charge-to-mass ratio as no greater than five per cent. The curve of Fig. 4 was got with aluminium as the metal which the alpha-rays bombarded. Curves were obtained in the same way with

carbon, with boron and with iron in place of the aluminium, and each had a peak at the proper situation for protons.

It seems, then, that the particles *are* protons. They are of the same kind, whatever the substance they come from; though the speed which

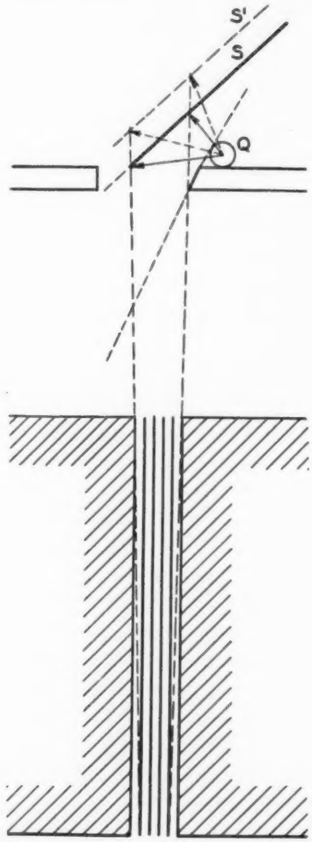


Fig. 3—Part of Stetter's apparatus for studying transmutation.

they have when expelled, and the plentifulness of the expulsions, varies notably from element to element. This suggests that they are constituents common to all elements, though the manner in which they are bound into the atomic structure differs from one to another. According to our knowledge of the astronomy of the atom, the nucleus is the only part where they can be. Moreover, though the masses of nuclei generally are not exactly integer multiples of the mass of the

proton, this is so nearly the rule as to suggest very forcibly that the major part of every nucleus consists of protons. All this strengthens the belief that in witnessing these flashes of "long-range" particles one is witnessing the signs of transmutation.

The next step, then, consists in finding which of the elements may be transmutable. I repeat that for the present, a strict assessment of the evidence permits us to proclaim a transmutation only when there are corpuscles of greater range than either the primary alpha-particles, or hydrogen nuclei which suffer elastic impacts. The condition, however, is not quite so harsh as I have intimated. If hydrogen atoms be

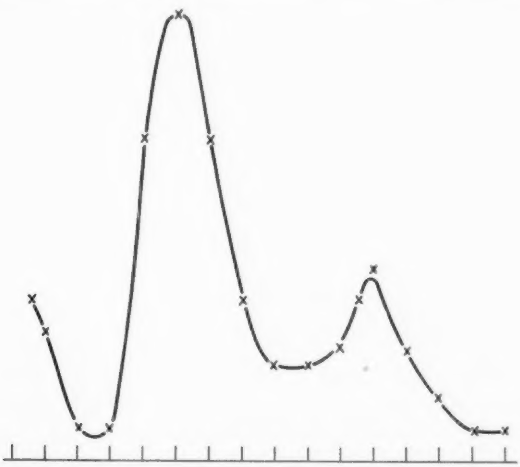


Fig. 4—Curve showing evidence that the particles emitted from aluminium bombarded by α-rays comprise protons and deflected alpha-particles (G. Stetter).

struck by alpha-particles, those and only those which are projected straight ahead have the full computed range; those which bounce off at an angle go less far; those which start off at 90° have no range at all which is to say, no elastic impact can send a nucleus off in the plane through the bombarded substance at right angles to the alpha-ray stream. Thus, if one stations the fluorescent screen somewhere in this plane, one may confidently count all of the scintillations as signs of transmutation, excepting such as may be due to primary alpha-particles deflected through 90° by kernels which they approach without disrupting them, or to hydrogen nuclei which after suffering elastic impacts got deflected. (The reader will have noticed in Fig. 3 that the angle between the paths of the α-particles to the metal foil and the paths of the protons from the foil is large, always more than 80°.)

Evidently it was the first of these possible causes of confusion which Rutherford feared the most, for in setting up his screen in such a place, he shielded it by absorbers sufficient to stop the primary particles, and counted the flashes which appeared in spite of this obstruction. The datum therefore is a count of corpuscles having ranges greater than seven cm. With a few of the lightest elements, the critical range is somewhat lower; for an alpha-particle, when deflected through 90° by a close approach to a nucleus not many times more massive than itself, loses an appreciable part of its speed in the deflection.

The Cavendish school examined many elements—including all of the nineteen lightest, the first nineteen of the periodic table—in their quest for transmutation. Under the bombardment of seven-centimeter alpha-rays, most of these nineteen emitted corpuscles which satisfied their strict criterion. The first four (hydrogen, helium, lithium and beryllium) did not; neither did the sixth nor the eighth (carbon and oxygen); all of the others, beginning with boron the fifth, and ending with potassium the nineteenth, appeared transmutable. No element beyond potassium ejected corpuscles with a range great enough to exceed those of the two other kinds, which Rutherford and his school were so anxious to exclude.

One is never long satisfied with the assertion that a certain effect does occur in certain cases, does not occur in others. Invariably the questions follow: in the cases where it happens, how much does it happen? in the cases where it is not observed, what is the least amount of it which could have been observed?

As a rule it is much more difficult to answer these questions, than merely to establish that with given means of observation either the effect is found or it is not. The study of transmutation is no exception to this rule. No one has ever set up a screen which surrounded the bombarded substance on *every* side, and therefore no one has counted the corpuscles which go off in *all* directions, nor even in a moderately great fraction of all directions. Screens have been set up in various directions from the piece of matter suffering transmutation, and the data, far from encouraging us to assume that the protons are fired off at random, indicate instead that more of them go off at inclinations of less than 90° to the beam (prolonged in the forward direction) of the alpha-rays, than at inclinations more than 90°,—more go "forward" than "backward." The distribution-in-angle, however, requires much further research.

In answering the questions, Rutherford, Chadwick and Ellis say *inter alia* that when one million alpha-particles of seven-centimeter

range are totally absorbed in nitrogen, some twenty of the rapid protons will probably emerge. This is the most efficacious degree of disruption which they claim. Aluminium follows in order of fragility, an equal bombardment producing eight of the fast-flying corpuscles instead of twenty. The least of the values given by the Cambridge school amounts to about one disruption per million alpha-particles; somewhat but not much smaller, I take it, is the least which they would deem observable, so that in their sense "immunity to transmutation" signifies something like this: when the substance is subjected to seven-centimeter alpha-rays, the number of protons coming forth with more-than-thirty-centimeter range is distinctly less than one per million thereof.

On the other hand, the physicists of the Vienna school have frequently maintained that transmutation is far less rare than those of the Cambridge school are willing to grant. Here, indeed, is one of the most famous controversies of modern physics. Vienna finds that most of the light elements, even carbon and oxygen, and even a metal so heavy as iron, yield scintillations, which are to be ascribed to protons ejected from nuclei; Cambridge holds to the list aforesaid. Where Cambridge admits scintillations, Vienna finds them several times more numerous. The contrast is accentuated by the fact that the Viennese scientists worked with alpha-particles of smaller energy than those at first employed by Rutherford, although the work of Pose, which I shall presently review, has destroyed what formerly seemed to be the natural assumption that the slower the alpha-corpuscles, the less must necessarily be their ability to transmute. The controversy was made peculiarly difficult to judge by the fact that for several years no one outside of these two schools essayed to enter the field. Eventually, however, several did; the researches of Bothe and Fränz, of Pose, and of Pawlowski, spoke for the lower efficiencies of transmutation believed in by Rutherford, rather than the higher ones accepted at Vienna. Many studies of scintillations, many comparisons of the scintillation method with the other methods, have resulted from this controversy, and will probably be regarded in the course of time as its enduring good. The latest announcements from Vienna indicate that the number of protons detected by the ionization methods is systematically less than the number of scintillations; and as these comparisons are still under way, I will leave the matter here, especially since the experiments which I am about to describe have superseded some of the earlier ones.

These new and striking experiments involve a more thoroughgoing study of such curves as appeared in Fig. 2: a study in which not merely

the end-point of the curve is located, but the entire shape is considered, the conditions of the experiment being so fixed as to make this shape significant. The experiments are, in fact, made upon the distribution-in-range of the ejected protons. Strata of gas or films of solid are interposed in the path of these particles, and the number which get through various thicknesses of these obstructions is carefully measured. The "air-equivalent" of the obstructions is separately measured, and so one is able to plot a curve of which the abscissa R is range in air at conventional pressure and temperature (760 mm. Hg and 0° C., in the figures which I show next) while the ordinate is the number of corpuscles having ranges greater than R.

This newest and sensational work was done by Pose at Halle. I show in Fig. 5 his sketch of his apparatus. In the evacuated chamber

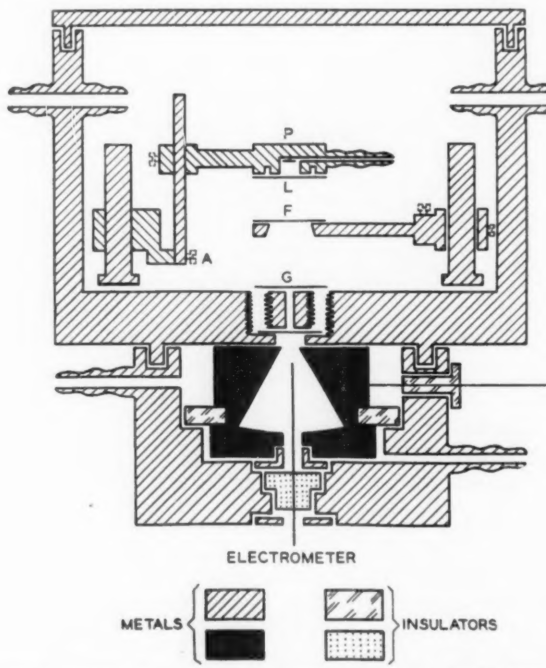


Fig. 5—Pose's apparatus for detecting transmutation of metals by an ionization method.

above, the button under the letter P carries the radioactive source—a layer of polonium, a substance of which the alpha-rays have a range smaller than those of which I have been speaking, 3.72 cm. only in air

at 0° C. and atmospheric pressure.[2] The foil of aluminium destined to be transmuted stands either at L or at F. If at L, it receives the full impact of the unretarded alpha-particles. If it is placed at F, one or more foils of gold are located at L (one at least *must* be set there, so as to prevent "contamination" of the chamber by atoms of radioactive substances escaping from the layer of polonium and wandering around) and these reduce the energy of the alpha-particles before they strike the leaf of aluminium. At G are placed the tenuous sheets of mica which retard or stop the protons, enabling the observer to plot the aforesaid curve (and incidentally protecting the vacuum within the upper chamber against the gas without). Below is the device for counting the protons which have succeeded in passing the mica sheets.

This device for detecting protons is not a fluorescent screen, but a conical chamber filled with carbon dioxide, in which the corpuscles engender thousands of ions as they cross. Between the walls of the cone and the wire which runs part way along its axis, there is a voltage sufficient to draw ninety per cent of the negative ions to the one, of the positive ions to the other. An electrometer connected to the wire gives a kick whenever a corpuscle passes through; the deflection is a measure of the total charge borne to the wire by ions of one sign, therefore of the total number of these. The kicks are not overly frequent; in cases mentioned by Pose they amounted to thirty or thereabouts per hour. They are not all equal; on the contrary, they range from almost imperceptible deflections (corresponding to 5000 ions or less) to a maximum which indicates seventy thousand. They are not all due to protons, for some are observed when the conical chamber is closed on all sides; these are ascribed to alpha-rays emanating from radioactive atoms which happen to be in the gas of the chamber or in the walls thereof; they are counted in "blank" experiments, and a number equal to theirs is deducted from the total number observed when the protons from the aluminium are coming in. It appears from the data that all of the corpuscles which produce more than 25000 ions apiece are of this undesired type, while most of those which cause the smaller kicks of the electrometer do actually come from the metal foil which is suffering transmutation.

Every detail of the set of curves next following (Fig. 6) is worth examining. They are curves of the sort which I defined above, except that the ordinate is not the actual number of protons observed, but the quotient of this number by that of the alpha-particles expressed in

[2] This is Pose's convention, to which I conform in the following pages; the range of alpha-rays from polonium is 3.92 cm. in air at 15° C. and atmospheric pressure; the other ranges mentioned in what follows should be increased in the same proportion if the reader wishes to hold to Rutherford's convention.

hundreds of millions—the number of protons received and detected in the conical chamber behind the absorption-foils of mica, per hundred million alpha-corpuscles impinging on the sheet of aluminium. This sheet was so thick that all of the impinging corpuscles were swallowed up in it. Thus, the uppermost of the curves relates to aluminium subjected to alpha-rays of all ranges from 3.72 centimeters down to nil; and the numbers attached to the others likewise stand for the *maximum*

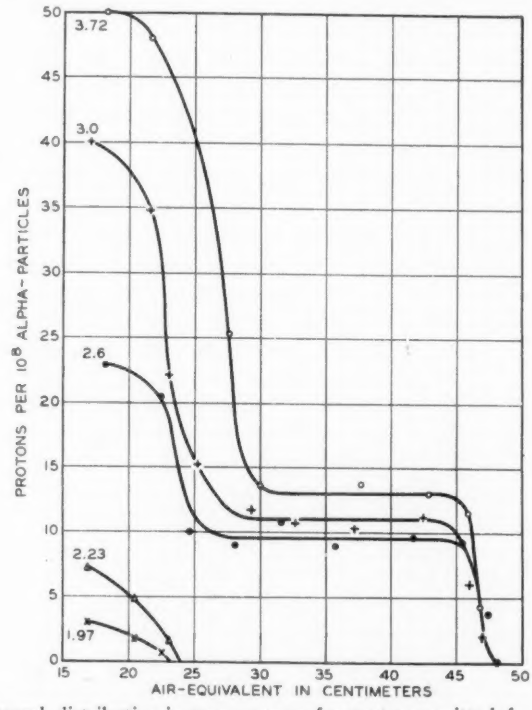


Fig. 6—Integral distribution-in-range curves for protons emitted from disrupted nuclei of aluminium atoms (H. Pose).

ranges represented among the particles as they approached the nuclei. By differentiating these curves one would get *distribution-in-range* curves for the protons. I will therefore refer to them, and to those of Figs. 2, 7 and 8, as "integral distribution-in-range curves."

Ignoring for the moment the left-hand half of the diagram, consider the other. The two lowest of the curves do not reach it: hence, alpha-particles of 2.23-centimeter range or less do not have the power of expelling protons with 30-centimeter range or greater. The other three

curves do enter it, and there is a very important feature of their trend: all are horizontal from 30 to 45, then all drop sharply to the axis at the same abscissa. Thus the piling-up of obstruction in the way of the protons does not stop them, so long as the air-equivalent is less than 45. But the adding of three further centimeters of air suffices to bar them all. It is as though the nuclei held protons in such a way, that if ejected at all they would automatically be ejected with velocities entailing ranges which lie within this narrow interval of 45 to 48; and alpha-particles acquired the power of setting off the mechanism by which these protons are ejected, when and only when their own range became as great as a critical value somewhere between 2.23 and 2.6.

Now travel back along the topmost curve into the left-hand half of the figure. The rise to the left of abscissa 30 suggests a second group of protons, having ranges slightly below this amount. But one notices, first, that the rise extends over an interval much wider than that of the steep sharp climbs at the right-hand ends of the curves; beginning at 30, it seems to be still going on at 18. This implies a broad distribution-in-range. One notices next that in the second curve, the corresponding rise begins at an abscissa somewhat smaller; in the third, at one which is smaller yet. Moreover, it is easy to draw a smooth curve through the starting-points of these three rises, which on being smoothly prolonged passes near to the points where the two remaining arcs in the lower left-hand corner ascend from the axis of abscissæ. All this suggests that in every one of these cases there are protons distributed over a wide interval of speeds, extending upward to a maximum which itself is greater, the higher the energy of the impinging alpha-rays.

Turn now to Fig. 7. Here we have five curves corresponding to five foils of aluminium, one face of each being exposed to alpha-rays of polonium with their full energy and undiminished range of 3.72 cm. The bottom curve relates to the thinnest foil, equivalent in thickness to 0.15 cm. of air. Actually, the distance between its two sides was the equivalent of 1 mm. of air, but many of the alpha-particles traversed it obliquely, so that the atoms of the foil were exposed to the blows of particles varying in range from 3.72 to 3.57 cm.; to this interval of ranges, therefore, the lowest curve refers. Similarly, the second curve from the bottom relates to a foil of air-equivalent 0.62 cm. for the most oblique of the particles, therefore to atoms bombarded by alpha-rays of ranges varying from 3.72 to 3.1; the other three, to sheets with the air-equivalents marked beside them. The difference between the second curve and the first is the effect of alpha-rays having ranges between 3.57 and 3.1. Thus, in going from one curve of Fig. 7 to the

next above it, one adds the effect of *slower* alpha-particles; whereas in Fig. 6, in going from one curve to the next above it, one adds the effect of *faster* projectiles.

The contrast between the lowest and the next-to-lowest curve of Fig. 7 is indeed amazing. So long as the impinging corpuscles are moving

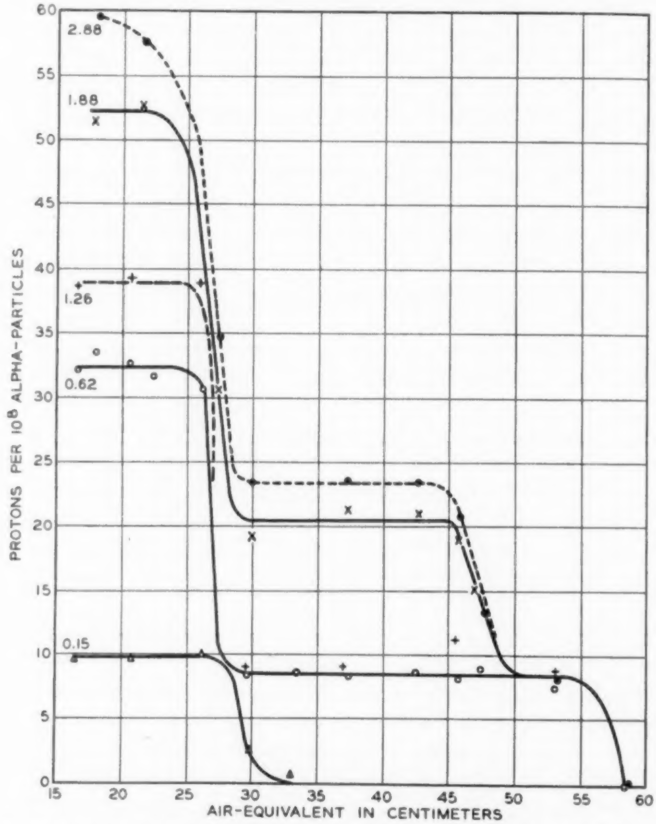


Fig. 7—More integral distribution-in-range curves for protons emitted from disrupted nuclei of aluminium atoms (H. Pose).

so fast that they still have 3.57 cm. of their range ahead of them, the protons which they eject are comparatively slow, with a maximum range of 30 or thereabouts; but *when they are slowed down* to some critical speed corresponding to some range between 3.57 and 3.1, *they acquire the power of dislodging extremely fast protons.* Here we have, in fact, another "group" in the sense of the previous pages: protons which

are released, if they are released at all, with a speed corresponding to a range in the neighborhood of 57.5 centimeters. The upper curves coalesce with this second-lowest over the descending arc at the right-hand end; the addition of slower alpha-rays to the bombarding stream does not increase the rate at which the members of this group are liberated; the power of dislodging them is confined to particles of a narrow interval of speeds.

The second-lowest and the middle curve are indistinguishable over the right-hand half of Fig. 7. In going from them to the second-highest curve, however, we meet another significant contrast. Another "group" of protons makes its appearance: its distinctive range is close to 47, it is evidently the very one [3] which was detected by scrutiny of Fig. 6. As alpha-particles slow down, they reach a critical speed at which they acquire the power of releasing this group. The corresponding critical range must be lower than (3.72 − 1.26) or 2.46 cm., for otherwise the rise near 47 would appear on the middle curve. It must be higher than (3.72 − 1.88) or 1.84 cm., or the rise would not appear on the second-highest curve.

Evidently there are both an upper and a lower critical range, R_2 and R_1, such that alpha-particles can cause the ejection of the protons of this group if and only if their ranges lie between R_1 and R_2. The curves of Fig. 7 "bracket" the upper limit of this interval, fixing R_2 between 1.84 and 2.46. Likewise the curves of Fig. 6 bracket the lower limit, locating it between 2.23 and 2.6. The interval must therefore lie between 2.23 and 2.46. By a more minute analysis of the curves, Pose locates it in the neighborhood of 2.42.

From the left-hand parts of the curves of Fig. 7, one makes the same deductions as from those of Fig. 6. Whatever their speed (within the scope of these experiments) alpha-particles possess the power of liberating protons with a wide distribution in range. The breadth of this distribution, i.e. the difference between the longest and the shortest ranges comprised within it, decreases with decreasing speed of the particles; so also does the longest range.

It appears, therefore, that there are two mechanisms of disruption. One seems to be controlled by the internal structure of the nucleus; the alpha-particle serves only to touch it off; it can be touched off, or actuated (to use a more dignified word) only by alpha-particles of a narrowly-delimited range of speeds; once it is actuated, it ejects a proton with a velocity strictly defined. The other accords more closely

[3] The group of range 57 cm. does not appear on the uppermost curve of Fig. 6, but Dr. Pose writes me that it was actually apparent in the data, and that he deducted it in order to make obvious the resemblance which exists between the right-hand ends of this and the next two curves when that group is disregarded.

with our idea of a smash. It can be achieved by alpha-particles of any speed, above (presumably) some minimum which in these experiments was not attained; the energy of the ejected proton increases with the energy of the projectile which brought about the crash.

Pose's investigation is one of several which in the last three years have been devoted to the ranges of the protons set free in transmutation. I chose to emphasize it because of the beauty and clearness of the curves, their long horizontal segments and sudden steep ascents which are the evidence for "groups" of protons; it is outstanding also because of the extent to which Pose controlled and varied the speeds of the alpha-particles. Other physicists, however, exposed a wide variety of elements to the bombardment of the alpha-rays, and observed the protons ejected at diverse angles to the direction along which the bombarding corpuscles came. These were Bothe and Fränz of the Reichsanstalt, and Chadwick, Constable and Pollard of the Cavendish Laboratory.

Bothe and Fränz attacked the light element boron; except for aluminium this is the element of which the transmutation has been most studied, for it seems to be much more liable to disruption by the relatively slow alpha-particles of polonium—those which have been used in most of the newer work—than the others commonly tested. Observing the protons projected more or less nearly straight ahead, these physicists plotted a curve comparable to those of Fig. 6; a curve, that is to say, whereof any ordinate represents the number of protons having ranges greater than the value given by the corresponding abscissa. Their curve had a horizontal segment—the first ever observed, so far as I am aware; for its historical interest I reproduce it here as the uppermost curve in Fig. 8. The sloping part to the right of that segment implies a group of protons having ranges between 33 and 23 cm.; the sloping part to the left, a distribution of ranges extending from some 20 cm. downwards. These inferences stand out more clearly from an inspection of the differential curve, which I exhibit here as Fig. 9.

Returning to this field of research (or, more probably, continuing in it uninterruptedly) Bothe and Fränz in 1930 published separate further papers. Bothe varied another factor—the angle between the direction along which the alpha-particles were coming, and that along which the particular protons which he observed were departing. In any one experiment, as we shall see, this angle varies over a wide range; one must specify its mean value, or some value near its mean; this I will denote by θ. Bothe, then, adjusted his apparatus so that to θ he could successively give several values between 0° and 116°; and some of the curves had a horizontal segment, or at least a flattish gently-

sloping section, adjoined on the right by a steeper descent to the axis of abscissæ—the sign of a "group" of protons comparatively fast. This descent occurred at smaller values of range, the greater the angle θ. This implied that the mean speed of the group depends on θ—the mechanism for ejecting these protons functions in such a way as to give less energy to those which fly off more obliquely. Pose extended his own researches on aluminium and obtained curves corresponding to the topmost of Fig. 6, for various values of θ ranging up to 135°. At this last cited angle, the ranges of the three groups had sunk from 57 and 48 and 28 to 45 and 38 and 20 respectively.[4]

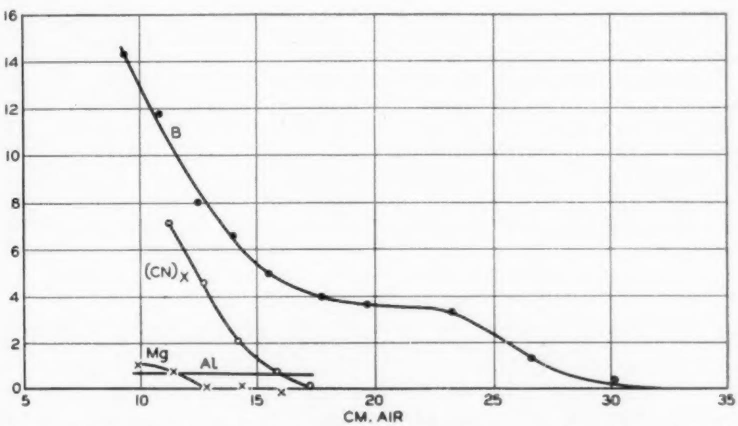


Fig. 8—Integral distribution-in-range curves for protons emitted from disrupted nuclei of boron and other atoms (W. Bothe and H. Fränz).

Whatever this fact may mean in regard to the mechanism, its practical consequence is clear. If the speeds of the protons depend on their direction of departure, then the interval over which these speeds are distributed for any given direction can be appreciated in its true narrowness (whatever that may be) solely by observing the protons which come off in that direction only. If in the actual experiment the paths of the alpha-particles falling upon the bombarded substance diverge over a wide solid angle, and if the paths of the protons which the counter or the fluorescent screen receives diverge likewise over a wide solid angle, then the sharpness of the groups must necessarily be masked. Now of course one would desire in any case to reduce these

[4] Before the discovery of groups, it had been observed at Cambridge that the maximum range of protons projected straight forward is greater than the maximum found among those projected almost straight backward: for boron the two values were 58 and 38, for aluminium 90 and 67 (in air at 760 mm. and 15° C.)

solid angles to the least practicable values. But in practice they *cannot* be reduced to low values, because the ejection of a proton is so infrequent an event that if one observed only those coming off within say a degree of a certain chosen direction, they would be altogether too few to be profitably observed during any reasonable period of time. The same thing would happen, if one used a beam of alpha-particles of similar narrowness; the impacts would be few because the impinging corpuscles were few. It is therefore to be feared that under the best of possible conditions, the horizontal segments in the curves will be shorter, the descents broader and smoother, than under ideal conditions they would be.[5]

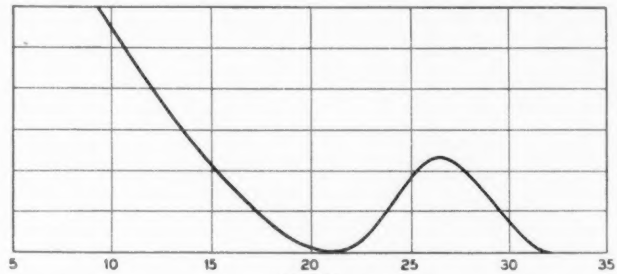


Fig. 9—Distribution-in-range curve obtained by differentiating the curve of Fig. 8 for boron (W. Bothe and H. Fränz).

Another result emerges from the experiments of Fränz, and those of the Cambridge school: the mean speeds of the groups apparently diminish with the speed of the α-particles. This is the effect which Pose observed with the slowest of the groups which he detected, not however with the faster and sharply-marked two; but from the experiments of the others, it seems to be the rule—not that the others tested all of the groups by varying the speed of the α-rays, far from it! but rather, for all which they did test, they found that sort of a dependence. Whether the constancy of speed of the groups which Pose studied is a peculiar feature of these, or his were the better experiments, I would not venture to say. At all events it is obvious that wherever this effect enters in, the natural sharpness of the groups is bound to be blurred by the differences in the speeds of the alpha-particles.

Another of Pose's discoveries—the remarkable fact that with aluminium, certain groups of protons are ejected only when the speeds of

[5] Whether the beauty of Pose's curves is to be ascribed to the smallness of the solid angles aforesaid is difficult to say. In one place he gives 0° and 58° as the range of values of θ, in another a somewhat smaller amount. Bothe says that in each of his observations the values of θ for 83% of the impinging α-particles lay within 15° of the mean—an interval of 30°. The others are not so definite.

the bombarding alpha-particles are kept within certain limited intervals—remains thus far unique. Nobody else has reported a similar observation.

As for the substances for which the distribution-in-range curve of the protons has been traced, for at least one speed of impinging alpha-particles and one value of the angle θ, they number seven. Of the German studies of boron and aluminium I have already written;

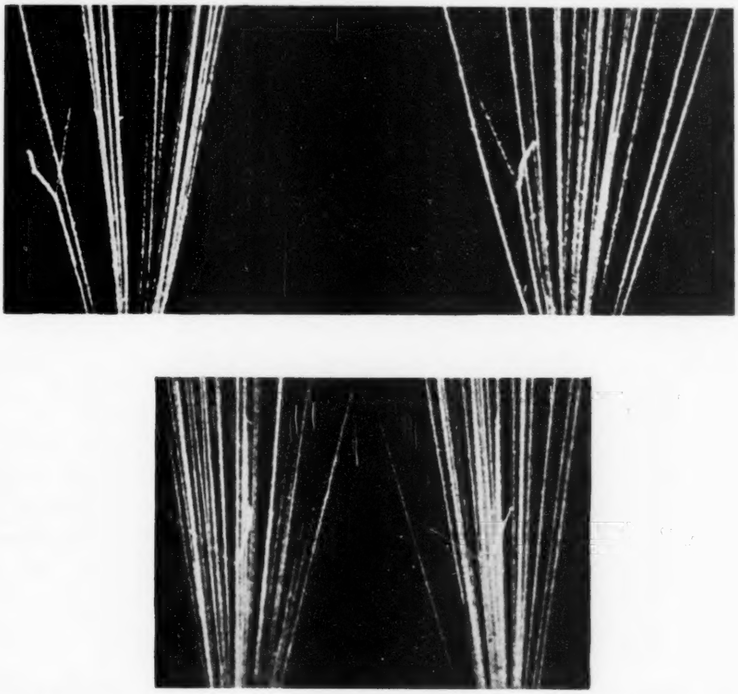


Fig. 10—Transmutation of a nitrogen atom attended by capture of the impinging alpha-particle (P. S. Blackett).

Chadwick and his colleagues also studied both, employing the alpha-particles of polonium with their pristine and with reduced speeds, and observing two groups with the former element, three (when the fastest particles were used) with the latter. With fluorine the Cambridge physicists observed three groups, with phosphorus one, with nitrogen a single group remarkably sharply defined. Sodium gave them a curve sloping smoothly downward without horizontal or even flattish segments. Lithium, carbon, oxygen, magnesium, silicon yielded under

these bombardments no protons at all, or at least none surely due to disintegration of nuclei. Yet Bothe and Fränz had observed protons issuing from magnesium, as one of the curves in Fig. 8 gives witness.

There remains the third method for detecting transmutation, the most beautiful and spectacular of all—the Wilson method, the one in which the trails of small charged particles are made visible by water condensing in droplets on the ions which the particles leave behind them as they tear through the gas. This seems a very inefficient

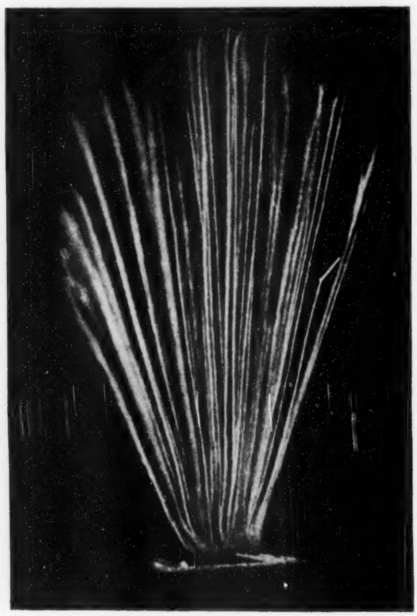


Fig. 11—Transmutation of a nitrogen atom attended by capture of the impinging alpha-particle (W. D. Harkins and A. E. Schuh).

scheme, considering how often one must photograph the trails of alpha-particles which do not effect a transmutation, before having the good fortune of finding on one's plate the record of one of the very rare alpha-particles which do. Inefficient it certainly is; nevertheless, by setting up apparatus which automatically repeats the experiment over and over and over again and automatically takes a new photograph at each repetition, one is able to assemble data enough to provide evidence of transmutation. Patience and perseverance are required, for there is one part of the process which cannot be delegated to a machine:

the personal inspection of the myriads of photographs, to locate those few which display "forked trails."

These forked trails were first studied by Blackett of Cambridge, bombarding nitrogen with very fast α-particles of 8.6 cm. range. Most of the few which he found are signs merely of elastic impacts: the alpha-particle has rebounded from an atom-nucleus leaving it intact, as one elastic ball rebounds from another; one of the two tines of the fork is the path of the recoiling nucleus, the other that of the rebounding alpha-particle. Nevertheless, among the trails of *two hundred and seventy thousand* alpha-rays of 8.6 cm. range,[6] Blackett found eight which were bifurcated in an evidently different way. Not, as he had expected, that there were three prongs to the fork instead of two. One would anticipate a long thin track for the proton (long because of its great range, thin because it produces fewer ions and therefore fewer droplets per unit length of its path), a short thick one for the alpha-ray after its impact, another short thick one for the recoiling residue of the nucleus. Actually in these eight cases there was a long thin track, undoubtedly that of the proton; and one, but only one, short heavy track. Harkins and two of his pupils, Shadduck first and later Schuh, made a similar search; chance was not so gracious to them as to Blackett; in the first research two forked trails were detected (not counting those resulting from elastic impact) among two hundred and fifty thousand; in the second, the same small number among an equal multitude.

This lack of a third prong to the fork probably means that the α-particle coalesces with the nucleus which it has just bereft of a proton, the solitary short track being the path of the resultant lately. This must be a nucleus of charge $+ 8e$; for the charge of the nitrogen nucleus is $+ 7e$, and to it has been added the charge $+ 2e$ of the alpha-particle, and from it has been deducted the charge $+ e$ of the proton. [As usual, e here stands for the magnitude of the fundamental electric charge, $4.77 \cdot 10^{-10}$ electrostatic unit.] Further, it must have a mass approximately equal to 17, on the familiar chemical scale on which the oxygen atom has mass 16; for the masses of nitrogen nucleus, alpha-particle and proton are approximately 14, 4 and 1 upon this scale. The ordinary atoms of oxygen have nuclear charge $+ 8e$ and mass 16. This new particle thus has the nuclear charge of an oxygen atom, not, however, its mass. It is consequently an "isotope" of ordinary oxygen.

[6] Actually, there were somewhat more than half as many additional trails due to α-rays of shorter range (5 cm.). The calculations mentioned in the next paragraph but one indicate, and practically prove, that all of the transmutations were performed by the faster rays.

To be so explicit about a particle, the existence of which is deduced from a set of a dozen forked alpha-particle trails which have one prong too few to the fork, may seem audacious. The evidence of the trails is, however, pretty strong.[7] Blackett measured with great accuracy the angles between the three trails, "stem" and "prongs" of the fork; to do this it was necessary to double the number of photographs, taking two simultaneous pictures from different directions every time the machine operated, so that by combining the two one could in effect "view" every fork in three dimensions. The two prongs and the stem always lie in one plane; and this is a necessary condition for conservation of momentum in a process in which the entire momentum of the impinging particle is shared by two and only two. If one could determine with perfect accuracy the speeds of the three corpuscles reasponsible for stem and prongs, one could tell whether or not the condition of conservation of momentum is obeyed, the masses of the corpuscles being put equal to 4 and 17 and 1, respectively. Or in other words, if the speeds of the corpuscles and their directions could be determined absolutely, one could compute by well-known formulæ the masses which they must have, in order to assure conservation of momentum. The speed of the alpha-particle is quite well-known; but for those of the two others, one is forced to depend on measurements of the lengths of their paths, combined with none-too-certain semi-empirical relations between their ranges and their velocities. Nevertheless, it was shown by Blackett that if the masses of the corpuscles responsible for the prongs of each fork are 17 and 1, the lengths and directions of their paths are such that so far as one can tell, momentum is conserved.[8]

Such is the present status of the art of transmutation. To the moment of this writing, it has proved so difficult that no one has been able to succeed in it except by using alpha-particles, nor to detect his success except by employing the delicate methods fit for perceiving individual fast-flying electrified corpuscles. Almost any day now, the first and perhaps also the second of these statements may cease to be true. Few scientific campaigns have ever enlisted so numerous, determined, energetic and powerful an array of talents and devices, as are now being

[7] Since these photographs were taken and interpreted, evidence has been found in band-spectra for the existence of an isotope of oxygen of atomic weight 17, very rare by comparison with the well-known one.

[8] Urey went further, and computed the mass of the residue to seven significant figures, utilizing the latest published values for the masses of alpha-particle and proton; he employed relativistic instead of Newtonian mechanics; in the former, the expression for momentum involves rest-mass and speed in such a way that conservation of momentum requires definite values for each, and these he calculated from Blackett's data (so, at least, I interpret his paper, but for the details I must refer the reader to it). The values which he gets range from 17.00504 to 17.00135 for the forked trails; these differences he believes to be real, inferring that the residual nucleus is left in different states by the different impacts.

devoted in the hope of imparting to free electrons and to protons such values of kinetic energy as heretofore only alpha-particles has possessed.

LITERATURE

Cambridge School:

For the work of Rutherford and his colleagues there is an authoritative summary in the book of Rutherford, Chadwick and Ellis "Radiations from Radioactive Substances" (MacMillan, 1930). Subsequently there has appeared the article by J. Chadwick, J. E. R. Constable and E. C. Pollard, *Proc. Roy. Soc.*, **A130,** 463–489 (1931). The paper of P. M. S. Blackett whence Fig. 10 was taken is in *Proc. Roy. Soc.*, **A107,** 349–360 (1925).

Vienna School:

The work of H. Pettersson and G. Kirsch and their associates at Vienna was described as of 1925 in the book of Pettersson and Kirsch, "Atomzertrümmerung" (Leipzig, 1926), which was followed three years later by a popular work of Pettersson alone, "Künstliche Verwandlung der Elemente" (Berlin, 1929). Among subsequent papers there are the following dealing directly with transmutation: R. Holoubek, *Zs. f. Physik,* **42,** 704–720 (1927); E. A. W. Schmidt, *ibid.*, 721–740; Kirsch and Pettersson, *Zs. f. Physik,* **51,** 669–695 (1928); G. Stetter, *Phys. Zs.*, **30,** 812 (1929). There are also numerous papers concerning the methods. Stetter's above-cited measurement of charge-to-mass ratio for protons is in *Zs. f. Physik,* **42,** 741–758 (1927).

Other Schools:

W. Bothe: *Zs. f. Physik,* **63,** 381–395 (1930). W. Bothe and H. Fränz: *Zs. f. Physik,* **43,** 456–465 (1927); **49,** 1–26 (1928).

H. Fränz: *Zs. f. Physik,* **63,** 370–380 (1930); *Phys. Zs.*, **30,** 810–812 (1929).

W. D. Harkins and R. W. Ryan, *J. Am. Chem. Soc.*, **45,** 2095–2107 (1923).

W. D. Harkins and H. A. Shadduck, *Proc. Nat. Acad. Sci.*, **2,** 707–714 (1926); *Zs. f. Physik,* **50,** 97–122 (1928).

W. D. Harkins and A. E. Schuh: *Phys. Rev.*, (2) **35,** 809–813 (1930).

C. Pawlowski: *C. R.*, **191,** 658–660 (20 October 1930).

H. Pose: *Phys. Zs.*, **30,** 780–782 (1929); **31,** 943–945 (1930); *Zs. f. Physik,* **60,** 156–167 (1930); **64,** 1–21 (1930); **67,** 194–206 (1931).

H. C. Urey: *Phys. Rev.* (2) **37,** 923–929 (1931).

Developments in Short-Wave Directive Antennas *

By E. BRUCE

Part 1 of this paper discusses the relative importance of the factors which limit the intelligibility of short-wave radio telephone communication. The more important of these factors are inherent set noise, external noise (static, etc.), and signal fading. The possibility of counteracting these limitations through antenna directivity is indicated.

Part 2 describes an antenna system which maintains a desirable degree of directivity throughout a broad continuous range of frequencies. The cost of this antenna is more favorable than that of many types of fixed frequency antennas of equal effectiveness.

BEFORE discussing specific antenna systems, it appears desirable to review the general problems of short-wave communication and to observe wherein antenna design can assist in overcoming existing circuit limitations. Accordingly, this paper is divided into two parts; the first will outline the requirements in the problem, and the second will be a description of an antenna system which has proved effective, despite its low cost of construction.

The writer's experience with antenna systems has been largely confined to the standpoint of reception, therefore, the following discussion will be largely on this basis. It will be apparent to the reader, however, that many of the features are likewise applicable to transmitting antenna installations.

PART 1. THE SHORT-WAVE PROBLEM

Radiotelephone Circuit Limitations

An analysis of the factors limiting the excellence of the output quality of a receiver governs the design of the entire radio circuit and associated equipment. Assuming well-designed apparatus throughout, we still encounter difficulties, especially at times of low signal strength, the more important of which are enumerated as follows:

(a) Inherent receiver noise.

(b) External noise (static, man-made noises, etc.).

(c) Signal fading.

The design of the receiving antenna system has an important bearing upon all three of these factors, brief explanations of which are given below.

* Presented before Sixth Annual Convention of the Inst. of Radio Engineers, June 6, 1931, Chicago, Illinois. Published in Proc. I. R. E., August, 1931.

(a) Receivers of very high gain characteristics are troubled with an inherent noise adequately described as a "hissing" sound. This may be due to several [1] causes such as shot-effect, etc. Much of this noise can be minimized through proper design, the methods of which are beyond the scope of this paper. Finally, however, an apparently irreducible minimum of noise is encountered, commonly referred to as [2] "Johnson" or circuit noise. This noise, under conditions of matched impedances, is so related to the circuit signal efficiency that the ratio of noise to signal cannot be appreciably altered except through somewhat impractical expedients such as lowering the absolute temperature of the circuit. All this tends to show that the designer of receivers must eventually rely upon his being able to increase the signal outputs from antennas to override the residual receiver noise difficulties on low field strength signals.

(b) Unpublished work, by a member of our laboratories,[3] has indicated that on many occasions short-wave static is highly directional. Interfering signals and electrical noises of human making are, of course, directional. It is quite evident that where the desired signal direction differs from that of the interference, receiving antenna directional discrimination is of immense importance.

(c) At times, remarkable reductions in short-wave fading have been achieved through extremely sharp directional characteristics of the receiving antenna. On the basis that certain types of fading are due to phase interference between multiple path signals of varying path length, it is reasonable to believe that where an angular difference exists between these paths, fading can be reduced by directivity which accepts only one of the paths. This, of course, assumes that the accepted path is stable in its direction. When this is not true, the reduction of fading through directivity becomes difficult.

The Relative Importance of the Various Circuit Limitations

The most serious hindrance to reliable, long-distance, short-wave communication is the great loss in signal fields which accompanies magnetic storms. Maintaining service under such conditions, develops into a battle against set noise and static. It is during these periods that effective receiving antennas are the most appreciated. The research worker on receiving antenna systems always welcomes such periods for his experimental work, since he knows well that under con-

[1] F. B. Llewellyn, "A study of noise in vacuum tubes and attached circuits," Proc. I. R. E., February, 1930.

[2] J. B. Johnson, "Thermal agitation of electricity in conductors," *Phys. Rev.*, **32**, 97, 1928.

[3] K. G. Jansky, Bell Telephone Laboratories.

ditions of strong signals, a simple antenna appears to perform as well as one considerably more elaborate and expensive.

Fig. 1 will assist in comparing the relative importance of set noise and static interference. The figure is not intended to be strictly accur-

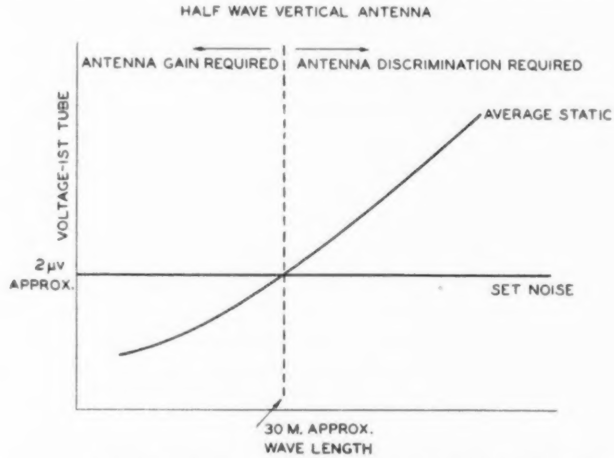


Fig. 1—Relative distribution of static and set noise with wave-length.

ate as to numerical values but will convey the idea of the principles involved. There is plotted as a function of wave-length, for an arbitrary location and season, the average static voltage level delivered to the first tube of a receiver by a half-wave, vertical antenna through its coupling circuits. Likewise, we have plotted the circuit noise delivered to this same tube as a function of wave-length. The fact that these curves intersect is of importance.

At wave-lengths considerably below the point of intersection, a weak signal falls into the level of the set noise. Increased signal output from the antenna is desirable to override this noise. It is evident that static reduction through directional discrimination is of little use in this region, therefore an antenna having directional properties but possessing no marked gain in output over a simple nondirectional antenna has no merit. At wave-lengths considerably above the point of intersection, static reduction through directivity is of utmost importance, while a gain in antenna output would be of little value if it meant a gain in static as well as in signal. It is interesting to observe, however, that a sufficient reduction of static through directivity would lower the whole static curve until it lay below the set noise curve. Such being the case, signal gain would again be required.

The above arguments are intended to show that, at the shorter wave-lengths, receiving antennas should be designed for a gain in signal output. At the long wave-lengths, directive discrimination in reception is the major requirement. In contrast to this, a transmitting antenna has no such wave-length eccentricities. Its purpose is always to lay down at the receiving point as great a field as possible. We must not forget, however, that the time is near when more attention should be paid to marked directive discrimination in transmitting antennas as a means of reducing interference between congested communication channels.

While set noise and static are at times important factors in limiting successful short-wave communication, fading practically always presents varying degrees of annoyance. It is really surprising how much fading can be tolerated without radically affecting speech intelligibility, but for services such as high-grade program transmission where naturalness is also important vast improvements are required; consequently much attention has been, and is being paid to this phase of the problem.

Increasing the Signal Output of Receiving Antennas

Under conditions of optimum output impedances, the magnitude of signal developed at the receiving antenna load is simply a function of the ratio of the effective induced voltage to the effective antenna resistance. The term effective induced voltage is used, as attention must be directed toward proper phasing, where the antenna dimensions are

Fig. 2—Effects of antenna directivity.

an appreciable part of a wave-length or more. Usually at short waves, the effective resistance is almost entirely the resistance equivalent of the reradiation losses. This resistance can be lowered through directivity, a simple example of which can be illustrated with the aid of Fig. 2.

If we can conceive of a point source of radiation at A, equipotential radiation surfaces would be spherical in shape and symmetrically disposed around A. The field intensity at point B would be unaffected if we had some means of avoiding radiation through the unshaded half

of the sphere, with a consequent saving of half of the radiated energy. If instead of saving this energy we added it to the shaded side, the energy available at B would be doubled. This is a simple explanation of the effect of directivity in the transmitting case. The receiving case is quite similar.

If the transmitter is at B, the energy available at A is diminished by reradiation losses. If we avoid reradiation through the unshaded half of the sphere, the radiation equivalent resistance is halved and the load energy will be doubled, after rematching the load to the antenna impedance.

With this knowledge of the usefulness of sharpened directivity, the designer is tempted to carry it to an extreme. The degree of directivity that may be beneficially attempted is, of course, limited by the variation in the apparent direction of wave arrival. For transatlantic, 16-meter signals over a daylight path, the horizontal plane angular variation, at New York has been [4] measured, by observing phase differences between spaced antennas, to be some 5 degrees or less, but apparently random throughout this range. Over a combination path of darkness and daylight, a horizontal angle variation considerably greater than this magnitude is frequently observed.

In the vertical plane, the variations in the apparent directions of arrival are considerable and also random. On rare occasions, angles as high as sixty degrees from the horizontal have been recorded. A sharp low angle antenna may well be expected to decrease in output as the angle of the wave direction becomes high.

Knowing that the interpretation of wave directions, by means of observed phase differences between spaced antennas, might be complicated if multiple waves of varying angles were present, two vertically polarized test antennas were built having optimum response at 27 degrees and at 6 degrees from the horizontal, respectively as shown in Fig. 3-A. These angles were experimentally obtained from airplane measurements. Fig. 3-B, which has been smoothed out for publication, is characteristic of about 80 per cent of the comparative data obtained on these two antennas, as measured by automatic signal recorders. Examination will show that, very frequently, the high angle antenna increases in output as the low angle antenna loses, or vice versa, indicating that the waves are varying in their vertical angle. Similar methods have also cross-checked the horizontal plane movements previously mentioned.

Where it is planned to design a single fixed antenna for a particular

[4] H. T. Friis, "Direction of propagation and fading of short waves," Proc. I. R. E., May, 1928.

service, the antenna should be sufficiently broad in its directivity to include most of the directional variations in signal arrival that may be encountered. In such cases, we have adopted the policy of simultane-

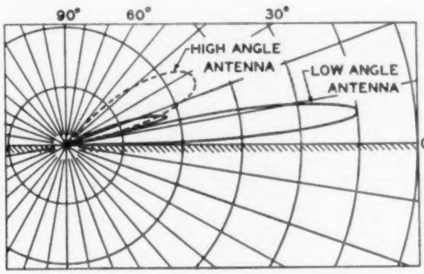


Fig. 3-A—Comparative directive diagrams of a high and a low angle antenna.

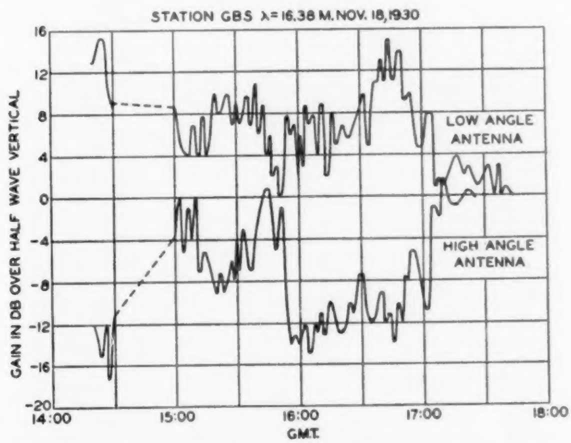


Fig. 3-B—Comparison of signal outputs of a high and a low angle antenna.

ously comparing the signal outputs of various size antennas through the measurements of automatic signal recorders over long periods. A photograph of one such signal recorder is shown in Fig. 4.

Several of our test antennas have proved to be too sharp. On occasions, their output exceeded that of any of the smaller, less directive antennas, but when averaged over long intervals of time, they proved to be deficient. At first, we tried to avoid putting too much weight on gain data obtained when signals were normally very strong but long

experience seems to show that wave direction variation has little correlation with the field strength of signals.

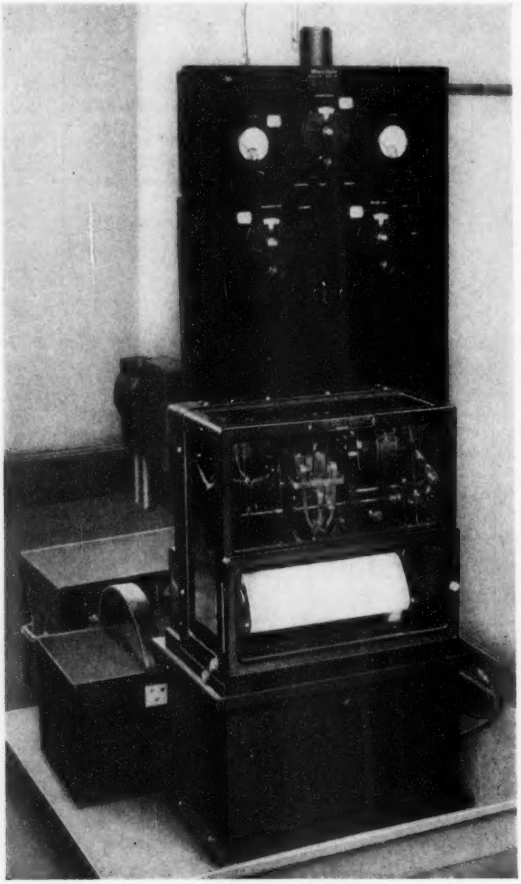


Fig. 4—An automatic signal recorder.

Static Reduction

Referring again to Fig. 2, assume that point A, receiving from B, is surrounded in all directions by static of uniform intensity. If A is made responsive only in the shaded directions, half of the static appears at first, to be eliminated, but we must remember that, by previous arguments, the static output from the shaded region is doubled; thus the

over-all static output is the same. For uniform distribution, the static output level is independent of the degree of directivity, provided that impedance matching between the load and the antenna is always maintained. We see, therefore, that the improvement in signal-to-static ratio in this case is the same as the signal improvement alone.

If static were always uniformly distributed about an antenna, the problems of signal gain and improvement in the signal-to-static ratio would be synonymous. The fact that short-wave static is usually highly directional puts an entirely different aspect on the problem. If, in Fig. 2, the static came from a direction included in the unshaded portion of the characteristic, the improvement in the signal-to-static ratio would be infinite. In a receiving antenna, therefore, emphasis must be placed on the deep suppression of response in other than the favored direction.

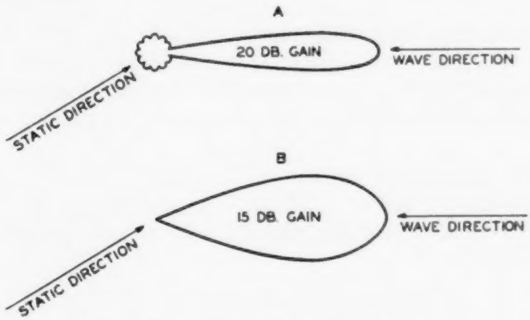


Fig. 5—A comparison of directive diagrams.

Fig. 5 is intended to illustrate the case described. The antenna characteristic 5-A, having a signal gain of 20 decibels over a nondirectional antenna, does not accomplish deep rejection in other directions. It follows, therefore, that the better discriminating characteristic 5-B would give a vastly better signal-to-static ratio, in spite of a smaller signal gain.

Fading Reduction

Many schemes for counteracting fading are in use and have been suggested. These include compensation for fading through automatic control of the receiver gain, the automatic selection of the best of several antennas, single side band with an unvarying locally supplied carrier, etc. All of these systems have merit, but are not a complete cure for the very prevalent selective type of fading, where several depressions may exist within a frequency band width of speech magnitude.

Under certain conditions, selective fading can be combatted through antenna directivity, but it is not without its difficulties in attainment. This is a direct attack on the multiple path source of the evil, eliminating a cause which makes fading selective with frequency. At times, very marked fading reduction has been obtained by this means.

Economics of Receiving Antennas

We have indicated briefly that the receiving antenna system has an important bearing upon all the major factors which are limitations in the present short-wave art. As long as these improvements can be effected in the receiving antenna system at a cost less than, for instance, a corresponding increase in transmitter power, concentration on the development of antenna design is well warranted.

One often hears the question whether one type of directive antenna is better than some other type. The answer usually depends on an economic comparison rather than an electrical one. The sharpness of directivity, the gain, etc., are determined by existing conditions. Numerous types of antennas can be designed to meet these specifications, therefore it is evident that the final selection is often based on over-all costs.

In Part 2 of this paper an antenna system will be discussed which is the result of an attempt to produce an effective antenna at a cost more favorable than the types we have been accustomed to use up to the present time.

PART 2. LONG WIRE ANTENNAS

Types of Directive Antennas

Directive methods, employing a finite number of spaced elements of specific phase and amplitude relations, have been known for a long time. Most of the more recent innovations, in this form of antenna, have pertained to the methods whereby, in their practical applications, these phases and amplitudes have been achieved. Considerable use has been made to date of such antennas, but they are quite expensive in their larger sizes and often their frequency range is very limited. As a result of these frequency restrictions, the radiotelephone receiving station at Netcong, N. J., employs ten [5] antennas, all differing in their design frequency but having the same favored direction toward England.

For some time, it has been appreciated that if it were possible to substitute a single directive antenna, having frequency characteristics sufficiently broad as to cover the above mentioned ten channels, a very

[5] A. A. Oswald, "Transoceanic telephone service—short wave equipment," *Bell Sys. Tech. Jour.*, April, 1930.

large economic saving could be effected. Development work was undertaken which has not only resulted in an antenna of considerable frequency latitude, but this new antenna structure is actually less expensive than a single, equally effective unit of the previous type. The remainder of this paper will be devoted to a discussion of various applications of this form of antenna.

Principles of "Tilted" Wire Antennas

The elementary principles underlying "tilted" wires can be explained more readily by presenting a physical picture, through the use of r-m-s vector representation, rather than through a more or less cumbersome mathematical treatment. The vector representations that follow are not rigorous but they serve to convey quickly the ideas under consideration and give results which are in sufficiently good accord with the complete mathematic analysis.

As we increase the length of a simple vertical antenna exposed to horizontally propagated waves, always rematching impedances by varying the load at its base, we obtain increases in the load power up to the point where the antenna wire length reaches one-half wave-length. The vector representation of this one-half wave-length case constitutes Fig. 6.

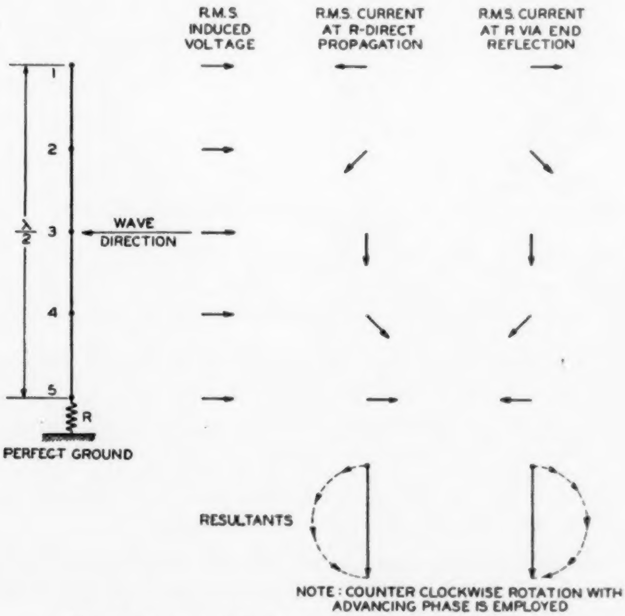


Fig. 6—Vector relations in a half-wave vertical antenna.

The first column of vectors represents the phase of the induced voltages, assumed to be lumped at points 1 to 5. The second column of vectors indicates the phase of the directly propagated currents arriving at R and due to each lumped voltage. The phase changes are due to the varying intervals of time required to traverse the intervening path. Likewise, the third column represents the current reaching R by way of the open-end reflection where a 180-degree phase change occurs. Summing up either column of current vectors, we trace a semicircumference and the resultant is a diameter. Had the antenna wire been slightly longer, the circumference would have been further closed and the resultant smaller. Fig. 7 illustrates an extreme case where the cur-

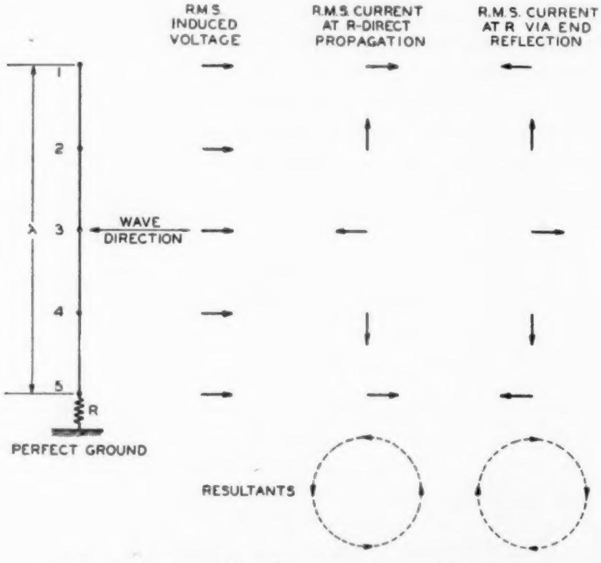


Fig. 7—Vector relations in a one-wave vertical antenna.

rents in R are zero for the vertical antenna length of one wave-length. Analyzing these vectors, we establish an important principle, as follows:

> The length of a straight antenna wire is an optimum value, for currents directly propagated to the load, when the elementary currents due to voltages induced in small lengths at the two wire extremities are opposite in phase at the load, provided that this does not also occur for intermediate points.

This statement has been restricted to the directly propagated currents since, in what follows, we shall, practically always, dissipate the cur-

rents propagated to the far end in appropriate terminating impedances. In many of the diagrams, the load currents which would arrive from open-end reflections have been included merely as of general interest.

The above stated principle permits us to remedy the null situation of Fig. 7 by tilting the wire as shown in Fig. 8. Notice that point 1 has been advanced into the wave propagation so that, at any given instant, point 1 is later in phase than for instance, point 5. The directly propagated currents of Fig. 8 trace a semicircumference and, therefore, the wire length [6] appears to be an optimum for the tilt selected.

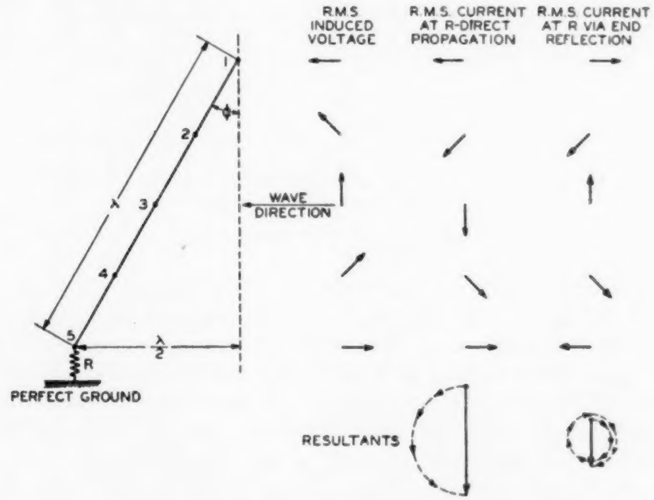


Fig. 8—Vector relations in a tilted wire antenna.

For any wire tilt angle, there exists a wire length which will trace a semicircumference similar to the above. This occurs when the tilt is such that the wire length is one-half wave-length longer than its projection upon the wave direction of propagation. Using appropriate tilt angles, as the wire length increases, output gains are achieved through increased effective induced voltage in the wire. Still further gain in output is available through the increasing directivity that is bound to result from the increasing dimensions.

One of the chief features of the tilted wire antenna is that in its

[6] For rigid accuracy in determining optimum dimensions, a small correction must be applied to these rules. This correction occurs in cases where, upon changing the wire tilt angle, the rate of change of induced voltage is comparable to the rate of change of load current as described above.

longer lengths it is effective over a broad range of frequencies. This is illustrated by Fig. 9 which is a plot of the wire length versus the tilt

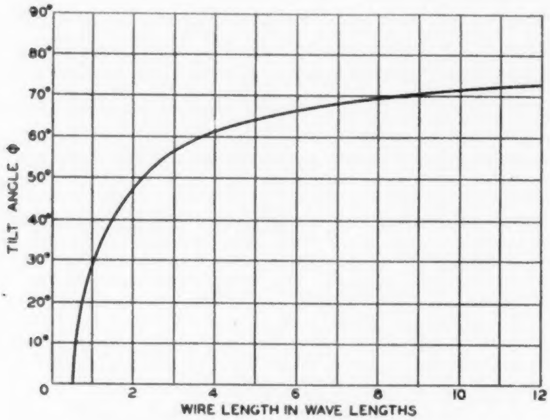


Fig. 9—Optimum tilt angle for long wires.

angle utilizing the above mentioned rules. For example, if the antenna were designed for a frequency such that the wire was ten wave-lengths long but it was used at another frequency where the wire length was only eight wave-lengths, Fig. 9 shows that the inaccuracy of tilt angle would be only about two degrees, which in most cases is inappreciable. As we shall see later, even this inaccuracy can be compensated by another wire in combination having an opposite trend.

Broad Frequency Range in Arrays

As is true for any antenna, the tilted wire may be used as an element in all the usual forms of arrays. Successful experimental antennas have been constructed consisting of a succession of tilted wires disposed in broadside relation, in the line of transmission and also stacked one above another. Some of these arrangements confine the effectiveness of the resulting antenna to a single frequency. Appreciating that one of the principle features of the tilted wire was its effectiveness over a broad frequency range, we have particularly stressed the development of those combinations of tilted wires which would not place restrictions on this frequency range. One such combination is discussed in the following section.

The Inverted V

The combination of two tilted wires to form the inverted V is shown in Fig. 10. The directional characteristics are appreciably improved

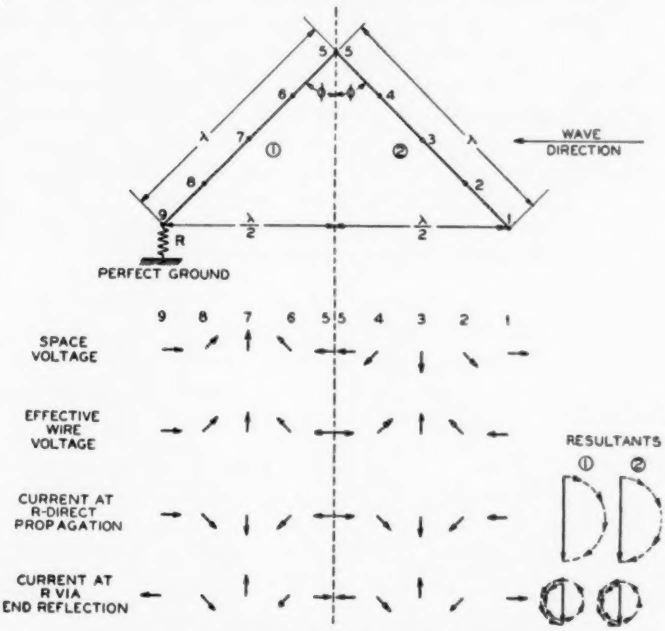


Fig. 10—Vector relations in an inverted V antenna.

with a consequent increase in signal output; also, the far end of the antenna becomes accessible for termination purposes, near the ground. These terminations will be discussed later. The inverted V requires no more supporting structure than the tilted wire, therefore its additional cost is very small where the land is available. Fig. 10 is a vector picture indicating that the two elements of the inverted V add in proper phase relation.

In connection with Fig. 9, it has been mentioned that the small inaccuries in tilt angle, due to departures from the design frequencies, can be counteracted by another wire in combination having an opposite trend. The inverted V of Fig. 10, is an example of one such possible arrangement. Since the tilt angle error is opposite in direction for each leg of the V, in combination, their optimum direction of response will remain unaltered. This will be illustrated by calculated directive diagrams which will be given later.

Asymmetrical Directivity Through Far End Terminations

Where it is desired to make an antenna responsive to signals in a given direction but to discriminate against signals in the opposite direction, reflector systems are often employed. These reflectors may be parasitic or they may be directly connected to the receiver through apparatus controlling their phase and amplitude relations. Our experience has shown that reflectors may be employed in connection with the type of antenna under consideration for the purpose of obtaining unilateral directivity. However, the use of reflectors restricts the possible frequency range, as they only function efficiently at specific spacings in relation to the wave-length used. For this reason, reflectors will not be discussed in this paper, although they are employed where a broad frequency range is not essential.

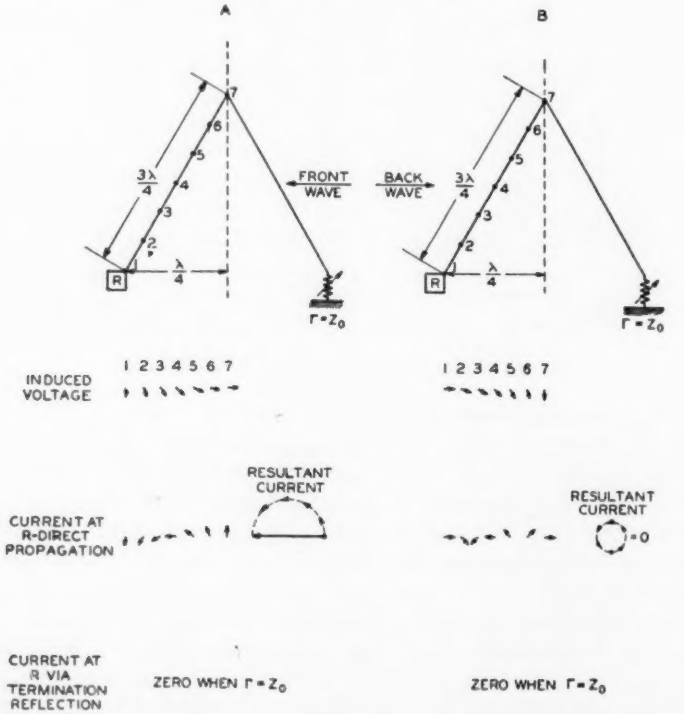


Fig. 11—Vector relations in an inverted V antenna—asymmetrical directivity.

Tilted wire antennas and their combinations are particularly adapted to obtaining directional asymmetry through proper terminations of the

end remote from the receiver. A simple example is illustrated in Fig. 11.

The end of the inverted V remote from the receiver R, in Fig. 11, is so terminated as to absorb signals without reflections. In other words, a termination equal to the antenna characteristic impedance is employed. Only the vectors for one leg of each of the inverted V's have been drawn, as the second leg is simply a reproduction of the first, and add directly thereto, after all phase relations have been determined.

In Fig. 11-A, a wave from the right produces elementary load currents which trace a semicircle, as previously discussed. Note that when the wave arrives from the left as in Fig. 11-B, the phase change is more rapid and a closed circle is traced making the resultant zero, thus we have achieved an infinite front-to-back ratio. It can be shown that this advantageous condition exists for tilted wires where the wire length of each element is an odd integral multiple, greater than one, of one-quarter wave-length, provided that the previously mentioned optimum tilt, in relation to the wave direction, is maintained.

At first glance, it might appear that the frequency range is restricted, since the above rule is limited to certain wire lengths expressed in wave-lengths. The most disadvantageous case exists when the wire length is an even integral multiple, greater than two, of one-quarter wave-length. Fig. 12 illustrates one such case, the wire being one

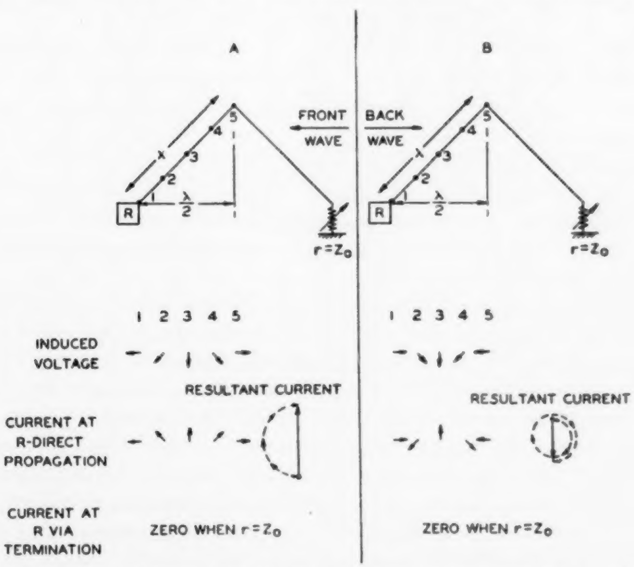


Fig. 12—Vector relations in an inverted V antenna—asymmetrical directivity.

wave-length long and at optimum tilt. It will be observed that the front-to-back ratio is not infinite but there still exists some directional discrimination, due to the fact that the back wave has resulted in the elementary currents tracing one and one-half rotations, thus obtaining partial cancellation. It is important to notice that longer wires would result in an increasing number of rotations and the resultant current of the back wave would become smaller and smaller as compared with the resultant of the front wave. This is a further argument for the use of long tilted wires. The calculated front-to-back ratios obtained with characteristic impedance terminations for various lengths of wires at optimum tilt are plotted in Fig. 13.

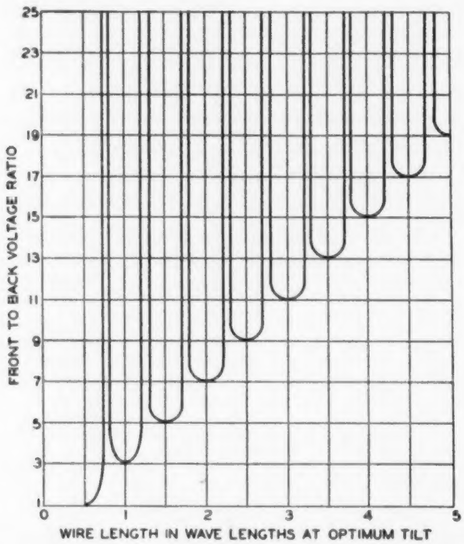


Fig. 13—Front-to-back ratios for characteristic impedance termination.

A very interesting feature about terminations is that, provided we are willing to make slight readjustments in their value, it is possible to obtain infinite front-to-back ratios at all frequencies within range. This is accomplished by cancelling the residue of back signal by means of a small reflection from the end termination obtained by departing slightly from the characteristic impedance adjustment. It can be shown that this results, for wires which are in length an even multiple, greater than two, of one-quarter wave-length, when the termination is the characteristic impedance times the cosine of the angle made by the wire with the direction of wave propagation.

For long wires, the above readjustment is very small. As an example, a ten-wave-length wire is properly tilted when it makes an angle with the direction of wave propagation whose cosine is 0.950. Thus, only a five per cent reduction in the termination from the characteristic impedance value will give an infinite front-to-back ratio.

In practice, we usually adjust a termination to a value which is a compromise between the above value and the characteristic impedance. This gives very favorable front-to-back ratios at any frequency within the range of the antenna, particularly in the case of long wires.

Theoretically infinite front-to-back ratios have been mentioned several times in the preceding discussion. It is an experimental fact that where very minute adjustments can be made in both the resistive and reactive components of the termination impedance, the front-to-back signal voltage ratio is only limited by the rigidity of the antenna elements in space. Voltage ratios in excess of 1000 to 1 are readily obtained, although such extremes are seldom warranted in practice. This deep depression can be "steered" through a considerable range of directions largely through changes in the reactive component of the termination impedance, the resistance alteration required being small. This permits a high degree of discrimination against many specific cases of interference in the rear quadrant of the antenna.

The Diamond-Shaped Antenna

In terminating inverted V antennas to ground, trouble has been experienced due to the instability of the ground contact resistance during varying weather conditions. In addition, the signal "pick-up" in the connecting leads was not always small compared with the antenna signal response in directions of antenna minima. These difficulties were avoided by terminating to the center point of a straight wire, substantially a half wave-length in total length, lying perpendicular to the favored wave direction.

As is well known, a quarter wave-length open-ended element appears to be a very low resistance when measured between its terminal and ground or another similar element. Two such low resistance quarter wave-length elements are effectively in parallel in the above arrangement and the center-tapped symmetry substantially balances out the effect of voltages induced in these elements.

Variations of the above type of artificial ground have been used in connection with inverted V antennas but, with few exceptions, they have required readjustments as the frequency was altered. A more satisfactory arrangement from several points of view is the double-V or diamond-shaped antenna shown in Fig. 14. This provides a bal-

anced arrangement eliminating the necessity of a "ground" connection; furthermore, it does not place any frequency limitation upon the system.

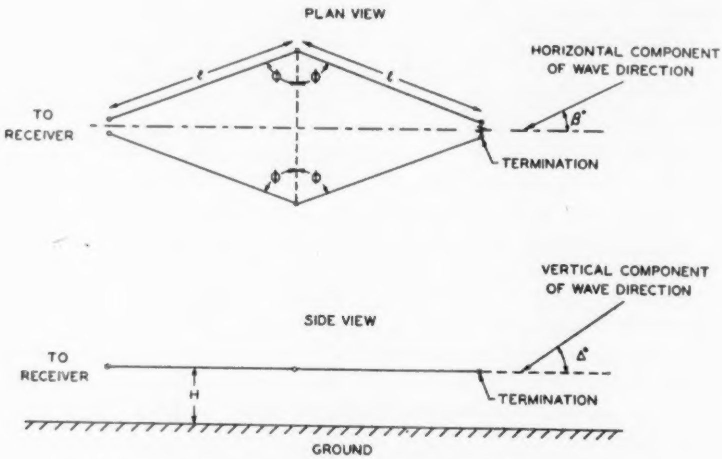


Fig. 14—The horizontal diamond-shaped antenna.

The antenna in Fig. 14 may be used with its plane either vertical or horizontal, being responsive, respectively, to vertically or horizontally polarized waves. It has found its greatest application in its horizontal form, however, due to reasons enumerated below.

(a) The supporting structure in its horizontal form is less costly, since only four relatively short poles are required.

(b) The inherent high angle directive characteristics of horizontal antennas discriminate against ignition, power, and other noises originating near the ground.

(c) The solid directive diagram of the diamond-shaped antenna is sharpest in the plane of the antenna. Since the direction of wave propagation is more stable in the horizontal plane, it is desirable to have the plane of the antenna horizontal.

(d) The directivity of the horizontal diamond-shaped antenna can be aimed, to some extent, at the most desirable vertical angle merely by altering the "tilt" angle ϕ of the antenna.

(e) The performance of the horizontal antenna is stable with varying weather conditions, since horizontally polarized waves are less affected than are the vertical by varying ground constants.

The use of the antenna horizontally, in the usual short-wave range, assumes that the strength of horizontally polarized waves are at least as great as are the vertically polarized components. Several observers have reported them more so, but the experience of the writer has been that there is little choice where horizontal and vertical antennas, having the same degree of directivity and optimum direction, are compared.

Up to this point in this paper, the attempt has been made to present simply a broad picture of some of the applications of long tilted wires to antenna design. It now seems worth while to give in somewhat more

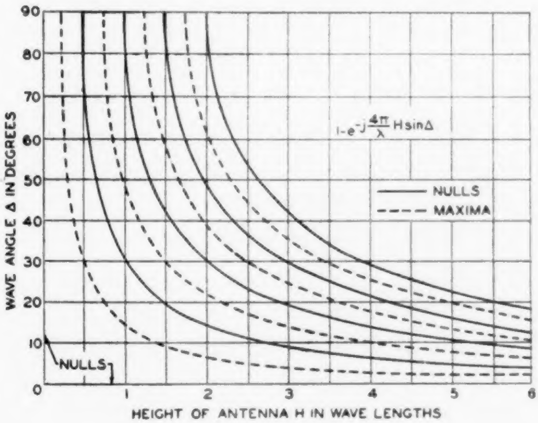


Fig. 15—Vertical plane design chart.

detail a sample of the design methods employed and the performance measurements on one typical form of antenna; accordingly a medium size horizontal diamond-shaped antenna has been selected.

The Horizontal Diamond-Shaped Antenna

In calculating the directive diagrams of the horizontal diamond-shaped antenna, the antenna wires have been assumed to be without resistance. As long as we are contented in knowing only the relative shape of the directive diagrams, this approximation is quite accurate and results in a tremendous simplification of the problem.

In all of the calculations, a perfect ground has been assumed. Fortunately, for horizontally polarized waves, variation in the ground constants do not radically affect either the amplitude or phase of the ground reflections, so that the following equations can be used as rough

approximations even where imperfect ground conditions are encountered.

Vertical Plane Directivity

The vertical plane directivity of the horizontal diamond-shaped antenna is determined by three factors, i.e., the length of each leg, the "tilt angle" and the height above ground.

For the cases where the element length is an integral multiple of a half wave-length and where the far end termination is the characteristic impedance multiplied by the sine of ϕ (see Fig. 14), the equation for the vertical plane directivity over perfect ground has been calculated to be,

$$I_R = k\left[1 - e^{-j4\pi H \sin \Delta/\lambda}\right]\left[\frac{1 + \cos \Delta}{1 - \sin^2 \phi \cos^2 \Delta}\right]\left[1 \pm e^{-j2\pi l \sin \phi \cos \Delta/\lambda}\right]^2 *$$

where, as shown in Fig. 14,

H = height above perfect ground in wave-lengths.
Δ = wave angle from horizontal in the vertical plane
ϕ = tilt angle of elements.
l = element length in wave-lengths.
k = proportionality factor.
I_R = receiver current.

It will be noted that neither the length nor the tilt angle appears in the first bracketed term. It can be shown that this factor appears as a multiplier for nearly any type of horizontal antenna, accordingly the location of nulls and maxima for this factor are separately plotted in Fig. 15.

In the same manner the nulls and maxima of the product of the second and third bracketed terms have been plotted in Fig. 16 for an element length of four wave-lengths.

The curves of Figs. 15 and 16 are design curves and their use can be illustrated by the following example: Measurements on the directions of wave arrival have indicated that the most usual directions are from 10 to 15 degrees above the horizontal. It is desired to construct a horizontal diamond-shaped antenna for this reception, employing four-wave-length elements. Fig. 15 indicates that the most economical pole height for 15 degrees is approximately one wave-length. Now referring to Fig. 16, we see that the largest tilt angle, to accomplish this, is about 65 degrees. It is always desirable to use the largest possible angle of tilt to obtain the use of the largest lobe of the directive diagram.

* In the third bracketed quantity use, in the $\pm$ sign, $-$ when l is an even integral multiple of $\lambda/2$ and $+$ when l is an odd integral multiple of $\lambda/2$.

Figs. 15 and 16 likewise give us the null points. These are seen to be 0, 30, and 90 degrees in Fig. 15 and 34, 57, 74, and 90 degrees in Fig. 16.

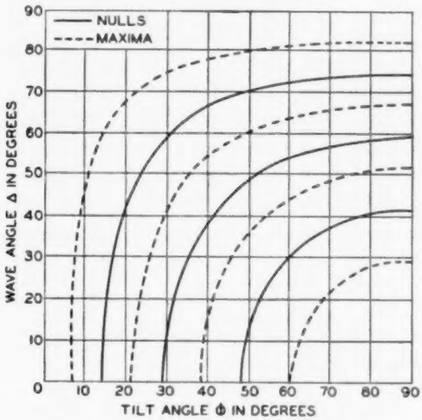


Fig. 16—Vertical plane design chart.

Using the above determined dimensions, the complete directive diagrams are calculated to determine whether a satisfactory result has been accomplished. Fig. 17 is the complete vertical plane diagram as

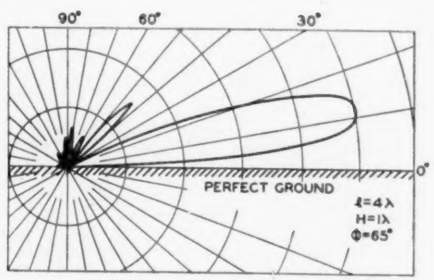


Fig. 17—Vertical plane directive diagram.

calculated from the previously given equation. Should some undesirably large minor lobe be present, it is often possible to suppress it by slightly changing one of the variables. A knowledge of the location of the null points, as given by Figs. 15 and 16, is a valuable guide in this accomplishment.

Horizontal Plane Directivity

Due to the cancellation effect of the reflections of horizontally polarized waves from a perfect ground, the horizontal plane diagram, for a horizontal antenna, is merely a point. The way to view directivity is properly in its solid form, but the calculations and plotted representations are somewhat laborious. The designer is in real need of knowing the horizontal width of the major lobe of the directional characteristic as would be seen from a plan view. This angular width, as measured between null points, is not altered by ground effects; therefore a useful simplification of the calculations may be had by ignoring the cancellation effect of the ground reflection. It should be pointed out that the amplitudes are slightly erroneous when this is done, but the null point locations are accurate. If this is done, we obtain the following equation:

$$I_R = k'\left[\frac{1+\cos\beta}{\cos^2\phi - \sin^2\beta}\right]\left[1 \pm e^{-j2\pi l\sin(\phi+\beta)/\lambda}\right]\cdot\left[1 \pm e^{-j2\pi l\sin(\phi-\beta)/\lambda}\right]^*$$

where, as shown in Fig. 14,

β = wave angle in horizontal plane.
ϕ = tilt angle of elements.
l = element length in wave-lengths.
k' = proportionality factor.
I_R = receiver current.

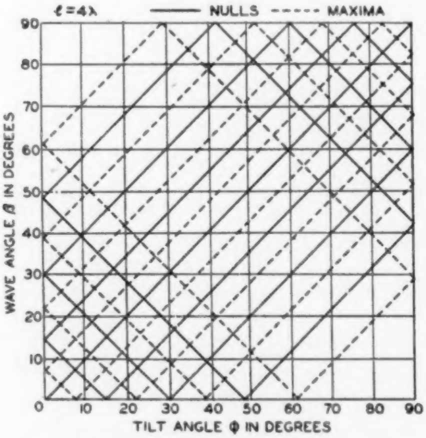


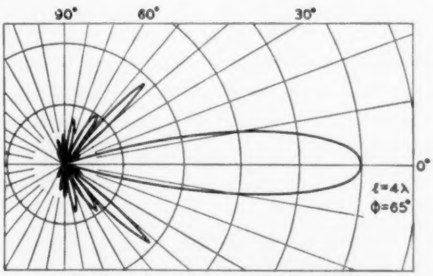
Fig. 18—Plan view design chart.

* In the second and third bracketed quantities use, in the $\pm$ sign, $-$ when l is an even integral multiple of $\lambda/2$ and $+$ when l is an odd integral multiple of $\lambda/2$.

Fig. 18 is a plot similar in character to that of Fig. 16, giving the location of nulls and maxima in the same manner. In our previous example, vertical plane considerations indicated that a tilt angle of 65 degrees was desirable. An examination of Fig. 18 gives a rapid estimate of the approximate plan view of the directive diagram and Fig. 19 is the more complete plan diagram for this tilt angle. It will be noted in Fig. 18 that the lines indicating factor maxima and minima frequently intersect. This property can be utilized for the suppression of particular minor lobes of the directive diagram by a proper selection of the tilt angle.

Frequency Range

Previously, it was stated that the V form of antenna counteracts the slight tendency for a change in optimum direction when the frequency



NOTE: GROUND CANCELLATION IS IGNORED

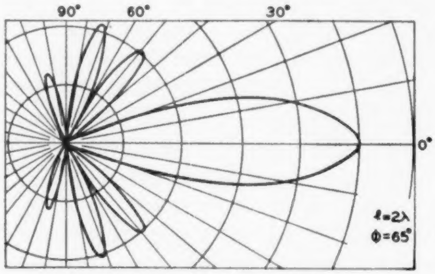
Fig. 19—Plan view directive diagram.

NOTE: GROUND CANCELLATION IS IGNORED

Fig. 20—Plan view directive diagram.

is altered. The correctness of this statement is verified in Figs. 19, 20, and 21. The linear dimensions and tilt angle were unaltered as the wave-length was varied over a two-to-one range. The optimum

direction is maintained although, as would be expected, the directivity becomes less sharp as the wave-length is increased in respect to the antenna dimensions.



NOTE: GROUND CANCELLATION IS IGNORED

Fig. 21—Plan view directive diagram.

Due to the variability of the wave directions in the vertical plane, this desirable direction is not well defined. As the wave-length is increased, a broadening characteristic counteracts the possibility of losing signal due to the optimum direction of the characteristic moving slightly upward.

Antenna Coupling Circuit

A two-wire transmission line has been used as the connecting link between the antenna and the coupling circuits at the receiver. With this arrangement, the circuits must be carefully balanced against vertical waves to obtain local noise reduction and to avoid reradiation losses from the transmission line. This is not difficult for a single frequency but if the coupling circuits are to maintain this balance for a range of frequencies, very careful designing of the coupling circuits is required.

The present practice is to place these coupling circuits in an elevated position directly at the antenna terminals to reduce the necessity for finical balancing adjustments. These circuits are connected to the receiver through a concentric pipe transmission line with its accompanying low loss, freedom from "pick-up," and substantial weatherproof construction. Multipeaked coupling circuits have been devised so that no readjustment is required over quite a frequency range.

Measured Performance

From the inception of our short-wave experience, we have been accustomed to compare the performance of antennas with a half-wave vertical antenna. The lower end of this standard of comparison is near

the ground and connected to a coupling circuit in such a manner that matched impedances are realized. Although the antenna under consideration is intended for the reception of horizontally polarized waves, the same vertical comparison standard has been maintained.

As previously mentioned, automatic signal recorders of the type shown in Fig. 4, are connected to each antenna. This recorder indicates an integrated average signal during each ten-second period, thus removing the wide amplitude excursions due to fading. It is an interesting fact that, although the instantaneous fading of two antennas may be different, the average signal over ten seconds usually has corresponding rises and falls in amplitude. This effect is so marked that any possible inaccuracies in the timing axis are readily detected, when comparing records. To promote accuracy in amplitude comparisons, only corresponding peaks or hollows of the curves are used. It is obvious that the employment of steep sides of curves would put a premium on very accurate timing. The relative timing of recorders is usually very good, as their synchronous motors are run by the same a-c power supply. The relative signal strength accuracy of the recorders is better than one db.

The antenna reported in the following data is an experimental antenna, at Holmdel, N. J., shown in the photograph of Fig. 22. This

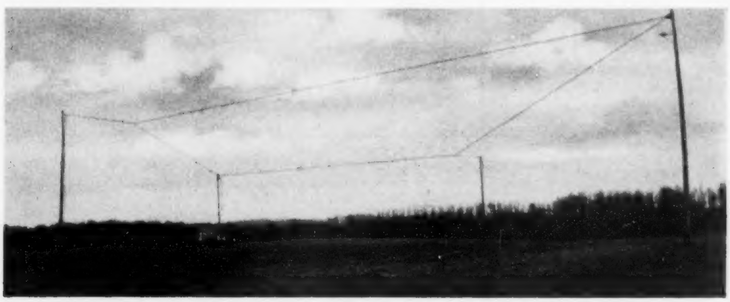


Fig. 22—An experimental horizontal diamond-shaped antenna.

picture illustrates the extreme simplicity of this type of antenna. The antenna dimensions are the same as those in the previously discussed directive diagrams when used at 16 meters. As has been said so many times before, the gain of the antenna over the standard may be expected to vary with the varying wave directions. The following data are the results of several hundred hours of tests, made at Holmdel, N. J., during the fall and winter months. Three different wave-lengths

were used with no alteration whatever in the antenna, its termination, or its transmission line coupling circuits. The standard of comparison, however, was always a half wave-length for the signal under test. It has been thought desirable to plot the gain data as the percentage of total time the antenna gain was above the indicated value in order to show the gain distribution with time. This summary of gains is given in Fig. 23.

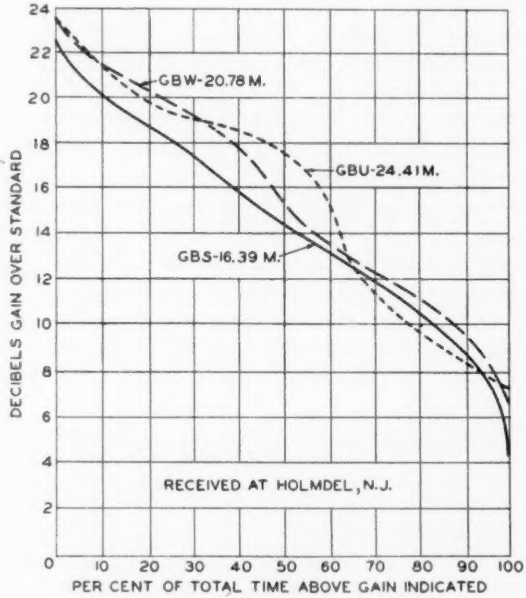


Fig. 23—Gain-time distribution curves.

I am indebted to a member [7] of our laboratories for an interesting variation which has been used in the application of this type of antenna to the transmitting problem. A simple terminating resistance is often undesirable in the transmitting case since it may be called upon to dissipate several kilowatts, in fact, that portion of the energy which would be radiated backward if no terminating resistance were employed. A long, two-wire iron transmission line shorted at the far end has been found to be one useful terminating load of the required dissipating ability.

The terminated diamond-shaped antenna possesses a broad impedance-frequency characteristic. This property may be augmented

[7] E. J. Sterba, Bell Telephone Laboratories.

by reducing the characteristic impedance of the antenna. One convenient scheme for reducing the impedance is to employ several conductors in parallel in each leg of the antenna. The characteristic impedance may in this manner be dropped to a value for which matching iron wire lines are readily constructed.

The terminating load which produces the most desirable impedance characteristic does not necessarily produce the best front-to-back ratio. In the transmitting case, however, the deep directed nulls required in reception, to eliminate interference of some particular station, are not necessary. It is sufficient to reduce by 10 or more decibels the field in the back directions. Thus the modified diamond-shaped antenna may be employed as a unidirectional transmitting array accepting power over a two-to-one frequency range.

In conclusion, I should like to point out that the work described in this paper was possible only through the assistance, coöperation, and advice of many people in the Bell System, to all of whom I render my sincere thanks. In particular, I wish to mention Messrs. A. C. Beck and L. R. Lowry who supervised the construction and did most of the testing of the experimental models. Mr. H. T. Friis, not only contributed many suggestions and constructive criticisms of the work, but took steps to have developed apparatus which was essential for the automatic measurement of received signal levels.

Abstracts of Technical Articles from Bell System Sources

Notes on Radio Transmission.[1] Clifford N. Anderson. Considerable data on radio transmission have been obtained the past few years in connection with the establishment and operation of various radio-telephone services by the Bell System. It is the purpose of these notes to present certain aspects of some of these data which may be of interest in the development of a general physical picture of radio transmission and in indicating the effects of disturbances accompanying storms in the earth's magnetic field.

The general results which are arrived at are:

1. Neglecting short time fading, the maximum field strengths received at a given point for frequencies up to at least 4 megacycles are in general agreement with those calculated by the inverse-distance law and the minimum field strengths (over-water transmission) are in approximate agreement with those calculated by the Austin-Cohen formula.

2. There appears to be a daylight absorption band in the neighborhood of 40 kilocycles (North Atlantic transmission) which reduces minimum daytime fields in that vicinity below the minimum limit given above.

3. The effect of solar disturbances is to increase the absorption to "sky wave" transmission throughout the entire radio-frequency spectrum generally and to reduce or eliminate the 40-kilocycle absorption band, thereby increasing daylight fields for transmission on frequencies in that vicinity.

Electrolytic Phenomena in Oxide Coated Filaments.[2] Joseph A. Becker. A critical survey of the literature shows that the current through the oxides in oxide coated filaments is carried by electrons, negative oxygen ions, and positive barium ions. The proportion of current carried by each depends upon the exact composition and method of preparation of the oxide coating, on the heat treatment and on previous electrolytic effects. Presumably the conductivity is greatly affected by barium and oxygen dispersed through the oxide. New experimental results show:

[1] *Proc. I. R. E.*, July, 1931.

[2] *Trans. Electrochemical Soc.*, Vol. LIX, 1931.

For a particular BaO + SrO filament, the conductivity C was . by

$$1.71 \times 10^4 \epsilon^{-\frac{1.73 \times 10^4}{T}} + 5.55 \times 10^{-3} \epsilon^{-\frac{0.62 \times 10^4}{T}}$$

2. The current is proportional to the voltage only so long as the current is small; otherwise the products of electrolysis alter the conductivity.

3. Polarization currents are caused by the Ba and O which are produced by electrolysis. These currents decrease rapidly even at temperatures near 500° K., thus showing that Ba and O diffuse at low temperatures.

Recent Developments in the Operation of Overseas Radio Telephone Service.[3] F. A. Cowan. This paper outlines the status of the present overseas radio telephone services from the United States, discusses the disturbing factors affecting each type of circuit, and outlines the reasons why short waves have come to be considered the probable medium for future extensions. The causes of lost circuit time to these services are given in their order of magnitude as: adverse atmospheric conditions, operating adjustments, radio interference, line and equipment troubles, and unclassified causes. Adverse atmospheric conditions have been partially overcome by the use of directive transmitting and receiving antennas and automatic gain devices on the radio receivers. These arrangements, however, do not eliminate the lost time caused by magnetic disturbances directly in the radio path or by the phenomena known as selective fading. Magnetic disturbances usually affect radio transmission over an appreciable period and a chart is included which shows the average manner in which they affected the available circuit time for the year 1930. The time required for operating adjustments which include such items as line-up and talking tests, changing wave-lengths, etc., will always be a factor but improvement will undoubtedly come with equipment development and experience. Line and equipment troubles are almost insignificant by comparison with the other causes of lost time and are made so by careful design and maintenance and the provision of spare units. A chart is included which shows for the month of August 1930 a comparison of lost circuit time, by causes, between the European and South American radio circuits. A chart is also included which shows the results of frequency measurements made over the month of August 1930 on the 21420 kc. transmitting frequency from the WLO transmitter at Lawrenceville, New Jersey. It is of interest to note that at no time during the month

[3] In abridged form, *Elec. Engg.*, July, 1931.

did the transmitter deviate from its assigned frequency by more than $\pm$ .01 per cent, whereas the limitation specified by the Federal Radio Commission is $\pm$ .05 per cent.

On the Art of Metallography.[4] FRANCIS F. LUCAS. Photomicrographs showing the highest degree of resolution and detail as yet obtainable with the high power microscope illustrate the paper.

Of particular interest is the new theory of the cause of fatigue failure in hardened steel presented by Dr. Lucas as due to the presence of minute cracks produced during the hardening process.

These quenching cracks average 25 atoms in width and 1000 atoms in length.

A complete description of the use and potential resolving ability of the high power microscope leads up to the art of metallography and its value in the industrial field.

Announcement is also made of the new metallurgical equipment by means of which can be achieved crisp, brilliant images at twice the present limits of useful magnification. The order of resolution will be improved and better optical and mechanical means will be at the disposal of the metallographer.

Some Physical Factors Affecting the Illusion in Sound Motion Pictures.[5] JOSEPH P. MAXFIELD. The advent of sound pictures brought the physicist and engineer face to face with problems which lie in the field of art as well as in the field of material things. A study of the physical factors which underlie art would probably be lengthy although it is conceivable that with sufficient knowledge of these physical factors it might be possible artificially to develop high-grade artistic sound pictures. It was felt, however, that more useful information of immediate applicability could be obtained by attempting to control, under the conditions of photography and recording, those factors which determine an observer's interpretation of what he sees and hears when observing a real event. The artist and director must be relied upon for the art in the production and the engineer or physicist is required to record and reproduce the scene in such a manner that the illusion in reproduction transmits to the audience the artistry produced by the actor.

This paper therefore describes the results of an empirical study of methods of controlling some of the factors available to the engineer in

[4] Presented at N. Y. mtg. of *Amer. Inst. of Mining and Metallurgical Engineers*, February, 1931. Published in *Heat Treating and Forging*, July and August issues, 1931.

[5] *Jour. Acous. Soc. Amer.*, July, 1931.

sound recording and photography in such a manner that a pleasing illusion of reality is created in the theater.

A Device for the Precise Measurement of High Frequencies.[6] F. A. Polkinghorn and A. A. Roetken. A description is given of equipment which has been constructed for the measurement of radio frequencies between 5000 and 30,000 kc. The equipment consists of a million-cycle quartz-crystal oscillator as a standard of frequency, means for producing harmonics and subharmonics of this frequency, and means for combining voltages of these known frequencies with a voltage whose frequency it is desired to measure so as to produce beat frequencies in successive stages, the beat frequency produced in each stage having one less digit than that in the preceding stage. A calibrated electric oscillator is used to measure the frequency of the last stage. An indicator gives the frequency of the unknown after a series of dial adjustments. The precision of a completed measurement is estimated at better than three parts in a million.

Radio Transmission Studies of the Upper Atmosphere.[7] J. P. Schafer and W. M. Goodall. In this paper are given a number of measurements which show time variations in the virtual height of the ionized regions of the upper atmosphere. These measurements were usually made simultaneously on two frequencies, 1604 kc. and 3088 kc. Single frequency data are also given. The following are the main points of interest presented.

(1) The data indicate the existence of two distinct ionized regions or layers. The changes in virtual height are sometimes very abrupt. The existence of the lower layer even at night is indicated by an occasional return to low virtual heights during this period.

(2) Experimental evidence has been found of large retardations in group velocity near the critical conditions for which the waves just penetrate the layer to the point of maximum ionization. (Fig. 1) Absorption is especially marked at such times.

(3) Except at these critical periods the records for the simultaneous transmissions show that the virtual heights of the upper layer are greater for the higher frequency than they are for the lower frequency. This statement would probably hold for the lower layer but no evidence on this point is presented.

(4) In the discussion several possible methods of two-layer formation are suggested, one of which involves the formation of negative ions in the region between the layers.

[6] *Proc. I. R. E.*, June, 1931.
[7] *Proc. I. R. E.*, August, 1931.

Theoretical and Practical Aspects of Directional Transmitting Systems.[8] E. J. STERBA. This paper discusses some of the more important principles involved in the development of the directional transmitting antennas at present employed in the Bell System short-wave facilities. The theoretical performance of directive arrays is presented by means of various curves which have been obtained by integrations based upon Poynting's theorem. The details of the mathematical derivations are omitted for the sake of brevity, but the general procedure and the resulting formulas have been placed in an appendix. Various practical problems encountered in the development are described. These include antenna tuning procedure, transmission line adjustments, and sleet melting facilities.

Nature of Stimulation at the Organ of Corti in the Light of Modern Physical Experimental Data.[9] R. L. WEGEL. The active prosecution of a program for the study of deafness has arrived at a point where a correct understanding of the mechanism of hearing may be utilized with profit. A "theory" should not be regarded as an academic description in terms of mathematical symbols of what is conceived to be a correct and final solution of the problem. It should be regarded as a necessary correlation of experimental data. The "correctness" of the theory should be judged by its utility and as long as it satisfies all demands made on it there is nothing "wrong" with it. The most that can be asked of any theory is that new experimental data, as it appears from time to time, will modify the conclusions only in quantitative detail but not in its broader qualitative aspects. The principal value of a theory is in the practical use that can be made of it, the value of it as an intellectual exercise being negligible.

The "theory" of hearing which apparently is in accord with all experimental data, whether it be anatomical, physiological or physical, is that which in its rudimentary form is known as the Helmholtz theory. Owing to the existence at present of a large quantity of precise data, particularly of a physical nature, this theory has undergone considerable advance since the time of Helmholtz.

Briefly, this theory ascribes the principal part of sound analysis to the mechanical properties of the end organ. In order to accept the essential points it is necessary to be agreed on a limited number of specific points:

1. If the basilar membrane vibrates with sufficient violence the hair cells in the superstructure of the organ of Corti are stimulated; and

[8] *Proc. I. R. E.*, July, 1931.
[9] Read before the *New York Academy of Medicine, Section of Otology*, Nov. 14, 1930.

further if, in response to a sound, the basilar membrane vibrates more violently in one place than in another, the stimulation of the nervous tissue is greatest where the vibration is most violent.

2. The basilar membrane does vibrate in response to sound and does so differently at different frequencies. It is easily shown by an elementary theory of mechanics that all bodies of whatever nature, whether solids, diaphragms, membranes, rods or bodies of fluid, behave in this fashion. Theoretically, it is possible to describe a body which vibrates the same at all frequencies, but such a body is never found experimentally. This leads to the conclusion that the basilar membrane where the nerve terminals are situated is quite capable of performing an analysis of a kind of sound.

3. The vibration of the basilar membrane resulting from sound is greatest at the proximal end for high frequencies and at the distal end for low frequencies. In order to arive at this conclusion Helmholtz depended on purely mechanical considerations, which for any one familiar with this type of philosophy is fairly satisfactory. Histological examination of ears known to have lowered acuity in certain frequency ranges have shown this to be the case.

4. In the normal ear there is only one spot which vibrates sensibly in response to one pure frequency in the cochlea. This thesis is quite well established by measurements on masking of one pure tone by another, in which case it is found that one sound masks another more effectively when the frequency of the second is nearer the first.

5. The only sensible functioning connections between nerve cells of the spiral ganglion, either direct or indirect, through branching of the peripheral axones at the organ of Corti, are confined to near neighbors. This thesis is also established by the physical data on masking.

6. The minimum detectable change of pitch corresponds to a shift along the basilar membrane of the vibrating spot for a distance equal to the space occupied by a definite number, approximately constant, of ganglion or hair cells.

With these points taken for granted it is possible to describe the mechanism of hearing in its broader aspects and to calculate to an approximation the actual position on the basilar membrane at which different frequencies stimulate it and to calculate also the extent of the stimulating spot for each frequency.

Automatic Power Plants for Telephone Offices.[10] R. L. YOUNG and R. L. LUNSFORD. The nature of power requirements for telephone offices is discussed, with emphasis on continuity of service. Auto-

[10] In abridged form in *Elec. Engg.*, June, 1931. In complete form, *Bell Tel. Sys. Monograph* B-561, May, 1931 and complete with discussion in *Trans. A. I. E. E.*, October, 1931.

matic controls are indicated because of their more exact performance, with consequent reduction in variations and in interruptions and their saving in maintenance, particularly with 24-hour operation demanded. Developments are traced, showing an increasing trend toward automatic regulation and control of main power supplies, ringing and other signaling energy sources.

The development of "unit type power plants" for telephone offices is discussed and information is given on a number of standardized plants which operate upon a full automatic or a semi-automatic basis. These furnish power supply for manual, toll, and telegraph central offices, for magneto offices and for manual and dial system private branch exchanges, also for small dial system central offices.

Favorable operating experience points the way toward further introduction of automatic devices which will place most telephone power plants, except those in the larger dial system offices, in a position to operate themselves over considerable intervals of time.

Contributions to this Issue

Austin Bailey, A.B., University of Kansas, 1915; Ph.D., Cornell University, 1920; Instructor in Physics, Cornell University, 1915–18; Signal Corps, U.S.A., 1918–19; Assistant Professor of Physics, University of Kansas, 1921–22; Department of Development and Research, American Telephone and Telegraph Company, 1922–. Dr. Bailey's work while with the American Telephone and Telegraph Company has been largely along the line of methods for making radio transmission measurements and of long wave radio problems.

Edmond Bruce, B.S., Massachusetts Institute of Technology, 1924. Radio service, U. S. Navy, 1917–19. Western Electric Company, 1924–25; Bell Telephone Laboratories, 1925–. Mr. Bruce has been engaged in the development of short-wave radio receivers and field-strength measuring equipment. More recently he has specialized in directive antenna systems for short-wave radio communication.

Karl K. Darrow, B.S., University of Chicago, 1911; University of Paris, 1911–12; University of Berlin, 1912; Ph.D., University of Chicago, 1917; Western Electric Company, 1917–25; Bell Telephone Laboratories, 1925–. Dr. Darrow has been engaged largely in writing on various fields of physics and the allied sciences. Some of his earler articles on Contemporary Physics form the nucleus of a book entitled "Introduction to Contemporary Physics."

Glenn D. Gillett. Studied at Pomona College; Harvard College, A.B., 1919; Harvard Engineering School, S.B. in E.E., 1921. Department of Development and Research, American Telephone and Telegraph Company, 1922–29, engaged in studies of radio field strength distribution and allied problems. Radio Development Group, Bell Telephone Laboratories, 1929–. Mr. Gillett has worked principally on common frequency broadcasting problems.

Thomas A. McCann, B.E.E., Ohio State University, 1925. Department of Development and Research, American Telephone and Telegraph Company, 1925–. Mr. McCann's work is chiefly in connection with printing telegraph systems.

W. B. Snow, A.B., Stanford University, 1923; E.E., 1925. Engineering Department, Western Electric Company, 1923–24. Acoustical research, Bell Telephone Laboratories, 1925–. Mr. Snow has been

engaged in articulation testing studies and investigations of speech and music quality.

A. L. Thuras, B.S., University of Minnesota, 1912; E.E., 1913. Laboratory assistant with U. S. Bureau of Standards, 1913–16. Graduate student in physics, Harvard, 1916–17. Bell Telephone Laboratories, 1920–. At the Laboratories, Mr. Thuras has worked on the study and development of electro-acoustic devices and instruments.

E. C. Wente, A.B., University of Michigan, 1911; S.B. in Electrical Engineering, Massachusetts Institute of Technology, 1914; Ph.D., Yale University, 1918. Engineering Department, Western Electric Company, 1914–16 and 1918–24; Bell Telephone Laboratories, 1924–. As Acoustical Research Engineer, Mr. Wente has worked principally on general acoustic problems and on the development of special types of acoustic devices.

R. I. Wilkinson, B.Sc., Iowa State College, 1924; Western Electric Company, 1920–21; American Telephone and Telegraph Company, Department of Development and Research, 1924–. Mr. Wilkinson has studied principally the application to telephone problems of the mathematical theory of probability, including sampling and statistical analysis.

The FARADAY *Centenary*

A Supplement to
The Bell System Technical Journal
October, 1931

M. Faraday

THE FARADAY CELEBRATION

ONE hundred years ago, Michael Faraday in the Laboratory of the Royal Institution, London, discovered the principle of electromagnetic induction. In this fundamental discovery lies the origin of the dynamo, the transformer, and the repeating coil—basic factors in the utilization of electricity for the purposes of man.

On the occasion of the centennial celebration of Faraday's discovery, Sir William Bragg gave the commemoration address in Queen's Hall, London, September 21, 1931. This address was broadcast in America, being transmitted across the ocean by radio.

An exhibition was opened in London on September 23rd, at which there were reproductions and illustrations of Faraday's actual experiments, prepared by the Royal Institution, together with a display of his chemical and electrochemical apparatus. There were also many exhibits by the great industries which exist today because of the practical application of Faraday's researches. Preceding the opening of the exhibition, General Smuts, recently elected President of the British Association for the Advancement of Science, gave an address.

On behalf of scientific and engineering societies in America, Dr. F. B. Jewett, President of the Bell Telephone Laboratories and Vice President of the American Telephone and Telegraph Company, speaking at Boston, Massachusetts, extended brief felicitations via transatlantic radio telephone and loud-speakers to those gathered in Albert Hall, London.

GREETINGS FROM SCIENTIFIC SOCIETIES OF THE UNITED STATES

To you, Mr. Chairman, to General Smuts, and to all those who have gathered in London today to commemorate the centenary of Michael Faraday's great discoveries in the opening of the Faraday Exhibition, I bring the greetings of the scientific societies and the men and women of science of the United States. In particular, I have been asked to convey to you the felicitations of the following societies which were invited to participate in the functions you have organized to evidence the world-wide appreciation of the debt we owe to a great man:

The National Academy of Sciences
The American Philosophical Society
The American Association for the Advancement of Science
The American Academy of Arts and Sciences of Boston
The New York Academy of Sciences
The American Mathematical Society
The American Physical Society
The American Chemical Society
The Franklin Institute
The American Institute of Electrical Engineers
The American Society of Civil Engineers
The American Society of Mechanical Engineers
The American Engineering Council
The Institute of Radio Engineers
The American Electrochemical Society
The United States Electrotechnical Committee
The Illuminating Engineering Society
The National Electric Light Association
The Association of Edison Illuminating Companies
The National Electrical Manufacturers Association
The American National Committee of the World Power Conference
The Association of Consulting Chemists and Chemical Engineers

Most, if not all, of these institutions are represented in London by their delegates. Neither they nor I can, however, express adequately the esteem in which Faraday and his achievements are held by tens of thousands of men who count themselves as his disciples.

Although I have formal authorization to speak only for my confreres in the United States, I feel quite safe in assuming in a degree to be the spokesman for men of science of whatever nationality. As such, I say to you of Britain that, although Faraday was of your blood, we of other lands yield you nothing in the measure of the respect and admiration in which we hold him. Go where you will in our institutions of learning, in the stately edifices we raise as homes for our scientific societies, or in the more prosaic housing of our scientific industrial establishments, and you will find always the evidence of our regard. For us he is ever a great simple man who enriched the world as few others have been privileged to enrich it.

In a way there is something peculiarly fitting in this tribute which I bring you and in the manner of its delivery. Involved in it is probably more of the fruit of all Faraday's works than can be encompassed in any other single happening in our modern world.

For me to sit here in Boston where Alexander Graham Bell made his great invention based on Faraday's discoveries and address you in London requires the application of something of all that men have learned in a hundred years in the fields to which Faraday opened the gates. Looking back across the years from the vantage point of our present achievements, it seems incredible that such vast things could have had such modest beginnings as those simple experiments of a simple kindly man. And yet, as we look forward through the eyes of a faith that has been trained to see distant things, it is clear that we have but embarked on the voyage.

While both you and I and those for whom I am speaking will long have passed to the great beyond, you may be assured that our descendants will join your descendants a hundred years hence when it comes time to commemorate another centenary of the man we are honoring today. For the present I can only reiterate that we in the far parts of the world are proud to have a spiritual part in your ceremonial.

FRANK B. JEWETT.

EXCERPT FROM FARADAY'S DIARY

FACSIMILE AND TRANSCRIPT OF THE PAGE RECORDING THE DISCOVERY OF ELECTRO-MAGNETIC INDUCTION *

Aug. 29th, 1831.

1. Experiments on the production of Electricity from Magnetism, etc. etc.

2. Have had an iron ring made (soft iron), iron round and $\frac{7}{8}$th inches thick and ring 6 inches in external diameter. Wound many coils of copper wire round, one half the coils being separated by twine and calico—there were 3 lengths of wire each about 24 feet long and they could be connected as one length or used as separate lengths. By trial with a trough each was insulated from the other. Will call this side of the ring A. On the other side but separated by an interval was wound wire in two pieces together amounting to about 60 feet in length, the direction being as with the former coils; this side call B.

3. Charged a battery of 10 pr. plates 4 inches square. Made the coil on B side one coil and connected its extremities by a copper wire passing to a distance and just over a magnetic needle (3 feet from iron ring). Then connected the ends of one of the pieces on A side with battery; immediately a sensible effect on needle. It oscillated and settled at last in original position. On *breaking* connection of A side with Battery, again a disturbance of the needle.

4. Made all the wires on A side one coil and sent current from battery through the whole. Effect on needle much stronger than before.

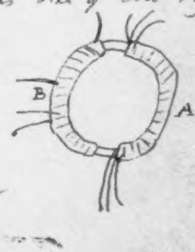
5. The effect on the needle then but a very small part of that which the wire communicating directly with the battery could produce.

* Courtesy of The Royal Institution of London, England.

Aug 29th 1831

Expts. on the production of Electricity from Magnetism &c

Have had an iron ring made (soft iron), iron round and 7/8 inches thick & ring 6 inches in external diameter. Wound many coils of copper wire round one half the coils being separated by twine & calico – there were 3 lengths of wire each about 24 feet long and they could be connected as one length or used as separate lengths. By trial with a trough each was insulated from the other. Will call this side of the ring A. On the other side but separated by an interval was wound wire in two pieces together amounting to about 60 feet in length the direction being as with the former coils this side call B.



3 Charged a battery of 10 pr. plates 4 inches square. Made the coil on B side one coil and connected its extremities by a copper wire passing to a distance and just over a magnetic needle (3 feet from iron ring). Then connected the ends of one of the pieces on A side with battery. Immediately a sensible effect on needle. It oscillated & settled at last in original position. On breaking connection of A side with Battery again a disturbance of the needle

4 Made all the wires on A side one coil and sent current from battery through the whole. Effect on needle much stronger than before

5 The effect on the needle then but a very small part of that which the wire communicating directly with the battery could produce

THE BELL SYSTEM TECHNICAL JOURNAL

A JOURNAL DEVOTED TO THE SCIENTIFIC AND ENGINEERING ASPECTS OF ELECTRICAL COMMUNICATION

TABLE OF CONTENTS

AND

INDEX

VOLUMES I—X

1922—1931

AMERICAN TELEPHONE AND TELEGRAPH COMPANY
NEW YORK

THE BELL SYSTEM TECHNICAL JOURNAL

Published quarterly by the
American Telephone and Telegraph Company
195 Broadway, New York, N. Y.

PRINTED IN U. S. A.

THE BELL SYSTEM TECHNICAL JOURNAL

TABLE OF CONTENTS

Volumes I–X (1922–1931)

Volume I, 1922

JULY, 1922

NOVEMBER, 1922

Volume II, 1923

JANUARY, 1923

APRIL, 1923

JULY, 1923

OCTOBER, 1923

Volume III, 1924

JANUARY, 1924

APRIL, 1924

JULY, 1924

OCTOBER, 1924

Volume IV, 1925

JANUARY, 1925

APRIL, 1925

JULY, 1925

October, 1925

Volume V, 1926

January, 1926

April, 1926

July, 1926

October, 1926

Volume VI, 1927

JANUARY, 1927

APRIL, 1927

July, 1927

October, 1927

Volume VII, 1928

JANUARY, 1928

APRIL, 1928

JULY, 1928

OCTOBER, 1928

Volume VIII, 1929

January, 1929

April, 1929

July, 1929

October, 1929

Volume IX, 1930

JANUARY, 1930

APRIL, 1930

JULY, 1930

OCTOBER, 1930

Volume X, 1931

JANUARY, 1931

APRIL, 1931

July, 1931

October, 1931

INDEX

Volumes I–X

(1922–1931)

A

B

D

E

G

H

I

J

K

L

M

Q

R

T

V

W